PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/54, 9/10, 2/21, 1/15, C12P 17/00, C12N 15/52, 15/53, C12P 19/62

A2

(11) International Publication Number: WO 97/23630

(43) International Publication Date: 3 July 1997 (03.07.97)

(21) International Application Number: PCT/US96/20238

(22) International Filing Date: 23 December 1996 (23.12.96)

(30) Priority Data:
08/576,626 21 December 1995 (21.12.95) US

(71) Applicant: ABBOTT LABORATORIES [US/US]; CHAD 0377/AP6D-2, 100 Abbott Park Road, Abbott Park, IL 60064-3500 (US).

(72) Inventors: SUMMERS, Richard, G., Jr.; W328 N3906 Lake County Drive, Nashotah, WI 53058 (US). KATZ, Leonard; 375 Plum Creek Drive, Wheeling, IL 60090 (US). DONADIO, Stefano; Lepetit Research Center, Via R. Lepetit, 34, I-21040 Gerenzano (IT). STAVER, Michael, J.; 162 Westerfield Place, Grayslake, IL 60030 (US).

(74) Agents: WEINSTOCK, Steven, F. et al.; Abbott Laboratories, D-377/AP6D-2, One Abbott Park Road, Abbott Park, IL 60064-3500 (US).

(81) Designated States: CA, JP, MX, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: POLYKETIDE-ASSOCIATED SUGAR BIOSYNTHESIS GENES

(57) Abstract

The present invention provides isolated polynucleotides from *Saccharomyces erythraea* that encode enzymes involded in the biosynthesis of polyketide-associated sugars. Methods of using the polynucleotides to produce novel glycosylation modified polyketides are also provided.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgystan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic			SE	Sweden
CG	Congo	KR	Republic of Korea	SG	Singapore
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LR	Liberia	SZ	Swaziland
CS	Czechoslovakia	LT	Lithuania	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	LV	Latvia	TJ	Tajikistan
DK	Denmark	MC	Monaco	TT	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	M	Madagascar	UG	Uganda
FI	Finland	ML	Mali	US	United States of America
FR	France	MN	Mongolia	UZ	Uzbekistan
GA	Gabon	MR	Mauritania	VN	Viet Nam

POLYKETIDE-ASSOCIATED SUGAR BIOSYNTHESIS GENES

This application claims the benefit of U.S. Serial No. 08/576,626 filed December 21, 1995, now pending.

Field of the Invention

The present invention relates to methods for directing the biosynthesis of specific polyketide analogs by genetic manipulation. In particular, sugar biosynthesis genes are manipulated to produce precise, novel glycosylation-modified macrolides of predicted structure.

Background of the Invention

Polyketides are a large class of natural products that includes many important antibiotic, antifungal, anticancer, and anti-helminthic compounds such as erythromycins, amphotericins, daunorubicins, and avermectins. Their synthesis proceeds by an ordered condensation of acyl esters to generate carbon chains of varying length, side chain, and reduction pattern that are differentially cyclized and subsequently modified to give the mature polyketides. For many polyketides, maturation includes the addition of one or more sugar residues to the cyclized carbon chain. The sugar residues are frequently critical to the biological activity of the mature polyketide.

Streptomyces and the closely related *Saccharopolyspora* genera are prodigious producers of polyketide metabolites. Because of the commercial significance of these compounds, a great amount of effort has been expended in the study of *Streptomyces* genetics. Consequently, much is known about *Streptomyces* and several cloning vectors exist for introducing DNA into these organisms.

Although many polyketides have been identified, there remains the need to obtain novel glycosylation modified (as defined herein) polyketide structures with enhanced properties. Current methods of obtaining such molecules include screening of biological samples and chemical modification of existing polyketides, both of which are costly and time consuming. Current screening methods are based on gross properties of the molecule, i.e. antibacterial, antifungal activity, etc., and both *a priori* knowledge of the structure of the molecules obtained or predetermination of enhanced properties are virtually impossible. Standard chemical modification of existing structures has been successfully employed, but is limited by the number of types of compounds obtainable. Furthermore, the poor yield of multistep chemical syntheses often limits the practicality of this approach. The following modifications to sugar residues bound to polyketides are particularly difficult or inefficient at the present time: change the stereochemistry of specific hydroxyl or methyl groups, change the oxidation state of specific hydroxyl groups, and deoxygenation of specific carbons. Accordingly, there exists a need to obtain molecules wherein such changes are specified and

performed which would represent an improvement in the technology to produce altered glycosylation-modified polyketide molecules with predicted structure.

The present invention overcomes these problems by providing the genetic sequence of sugar biosynthesis genes involved in the biosynthesis of polyketide-associated sugars.

Summary of the Invention

In one aspect, the present invention provides an isolated single or double stranded polynucleotide, typically DNA, having a nucleotide sequence which comprises (a) a nucleotide sequence selected from the group consisting of (i) the sense sequence of FIG. 4A (SEQ ID NO:1) from about nucleotide position 54 to about nucleotide position 1136; (ii) the sense sequence of SEQ ID NO:1 from about nucleotide position 1147 to about nucleotide position 2412; (iii) the sense sequence of SEQ ID NO:1 from about nucleotide position 2409 to about nucleotide position 3410 ; (iv) the sense sequence of FIG. 4B (SEQ ID NO:2) from about nucleotide position 80 to about nucleotide position 1048; (v) the sense sequence of SEQ ID NO:2 from about nucleotide position 1048 to about nucleotide position 2295; (vi) the sense sequence of SEQ ID NO:2 from about nucleotide position 2348 to about nucleotide position 3061; (vii) the sense sequence of SEQ ID NO:2 from about nucleotide position 3214 to about nucleotide position 4677; (viii) the sense sequence of SEQ ID NO:2 from about nucleotide position 4674 to about nucleotide position 5879; (ix) the sense sequence of SEQ ID NO:2 from about nucleotide position 5917 to about nucleotide position 7386; and (x) the sense sequence of SEQ ID NO:2 from about nucleotide position 7415 to about nucleotide position 7996; (b) sequences complementary to the sequences of (a); (c) sequences that, on expression, encode a polypeptide encoded by the sequences of (a); and (d) analogous sequences that hybridize under stringent conditions to the sequences of (a) and (b). A preferred molecule is a DNA molecule. In another embodiment, the polynucleotide is an RNA molecule.

In another embodiment, a DNA molecule of the present invention is contained in an expression vector. The expression vector preferably further comprises an enhancer-promoter operatively linked to the polynucleotide. In a preferred embodiment, the DNA molecule in the vector is one of the preferred sequences mentioned above. In an especially preferred embodiment, the DNA molecule in the vector is the sequence of SEQ ID NO:2 from about nucleotide position 80 to about nucleotide position 1048.

The present invention still further provides for a host cell transformed with a polynucleotide or expression vector of this invention. Preferably, the host cell is a bacterial cell selected from the group consisting of *Saccharopolyspora spp.*, *Streptomyces spp.* and *E. coli*.

The present invention also provides methods to produce novel glycosylation modified

polyketide structures by designing and introducing specified changes in the DNA governing the synthesis and attachment of sugar residues to polyketides. According to one method, the biosynthesis of specific glycosylation-modified polyketides is accomplished by genetic manipulation of a polyketide-producing microorganism comprising the steps of isolating a sugar biosynthesis gene-containing DNA sequence from those described above; identifying within the gene-containing DNA sequence one or more DNA fragments responsible for the biosynthesis of a polyketide-associated sugar or its attachment to the polyketide; creating one or more specified changes into the DNA fragment or fragments, thereby resulting in an altered DNA sequence; introducing the altered DNA sequence into a polyketide-producing microorganism to replace the original sequence whereby the altered DNA sequence, when translated, results in altered enzymatic activity capable of effecting the production of the specific glycosylation-modified polyketide; growing a culture of the altered polyketide-producing microorganism under conditions suitable for the formation of the specific glycosylation-modified polyketide; and isolating said specific glycosylation-modified polyketide from the culture.

In a second method the biosynthesis of specific glycosylation-modified polyketides is accomplished by isolating a sugar biosynthesis gene-containing DNA sequence from from those described above; identifying within the gene-containing DNA sequence one or more DNA fragments responsible for the biosynthesis of a polyketide-associated sugar or its attachment to the polyketide; reversing the strand orientation of the DNA fragment or fragments, thereby resulting in an altered DNA sequence which, when transcribed, results in production of an antisense mRNA; introducing the altered DNA sequence into a polyketide-producing microorganism having an mRNA capable of binding to the antisense mRNA which results in altered enzymatic activity capable of effecting the production of the specific glycosylation-modified polyketide; growing a culture of the altered polyketide-producing microorganism under conditions suitable for the formation of the specific glycosylation-modified polyketide; and isolating the specific glycosylation-modified polyketide from the culture.

In a third method the biosynthesis of specific glycosylation-modified polyketides is accomplished by isolating a sugar biosynthesis gene-containing DNA sequence from from those described above; identifying within the gene-containing DNA sequence one or more DNA fragments responsible for the biosynthesis of a polyketide-associated sugar or its attachment to the polyketide; introducing the DNA fragment or fragments into a polyketide-producing microorganism whereupon transcription and translation of the DNA fragment or fragments generate an altered polyketide-producing microorganism that is capable of producing the specific glycosylation-modified polyketide; growing a culture of the polyketide-producing microorganism containing the DNA fragment or fragments under

conditions suitable for the formation of the specific glycosylation-modified polyketide; and isolating the specific glycosylation-modified polyketide from the culture.

Preferably, the sugar biosynthesis gene-containing DNA sequence of the processes described above comprises genes which encode an enzymatic activity involved in the biosynthesis of L-mycarose and/or D-desosamine. More preferably, the sugar biosynthesis gene-containing DNA sequence comprises the sequence of SEQ ID NO:2 from about nucleotide position 80 to about nucleotide position 1048.

The present invention is especially useful in manipulating sugar biosynthesis genes from *Streptomyces* and *Saccharopolyspora*, organisms that provide over one-half of the clinically useful antibiotics.

Brief Description of the Drawings

FIG. 1A illustrates the organization of the erythromycin biosynthetic gene cluster and the genetic designations of the biosynthetic genes; FIG. 1B illustrates an abbreviated erythromycin biosynthetic scheme that broadly associates the biosynthetic genes with their role in erythromycin biosynthesis. Seven *eryB* genes, *eryBI - eryBVII*, are responsible for the biosynthesis of L-mycarose or its attachment to the erythronolide B ring, and six *eryC* genes, *eryCI - eryCVI*, are responsible for the biosynthesis of D-desosamine or its attachment to 3-α-mycarosylerythronolide B. The dashed arrows indicate that the pathway through erythromycin B is not the principal natural biosynthetic route to erythromycin A.

FIG. 2 illustrates the proposed scheme for the biosynthesis of L-mycarose and the *eryB* genes responsible for the specific steps.

FIG. 3 illustrates the proposed scheme for the biosynthesis of D-desosamine and the *eryC* genes responsible for the specific steps.

FIG. 4A(1-4) illustrates the nucleotide sequence (SEQ ID NO:1) of the sugar biosynthesis genes *eryCII* (coordinates 54-1136), *eryCIII* (coordinates 1147-2412), and *eryBII* (coordinates 2409-3410), with corresponding translation of the open reading frames (SEQ ID NO:3, SEQ ID NO:4 and SEQ ID NO:5 respectively). Standard one letter codes for the amino acids appear beneath their respective nucleic acid codons as described herein.

FIG. 4B(1-9) illustrates the nucleotide sequence (SEQ ID NO:2) of the sugar biosynthesis genes *eryBIV* (coordinates 80-1048), *eryBV* (coordinates 1048-2295), *eryCVI* (coordinates 2348-3061), *eryBVI* (coordinates 3214-4677), *eryCIV* (coordinates 4674-5879), *eryCV* (coordinates 5917-7386), and *eryBVII* (coordinates 7415-7996) with corresponding

translation of the putative open reading frames (SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11 and SEQ ID NO:12 respectively). Standard one letter codes for the amino acids appear beneath their respective nucleic acid codons as described herein.

FIG. 5A illustrates the amino acid sequence identity between the sugar biosynthesis enzyme encoded by the *eryBIV* gene of *Sac. erythraea* (SEQ ID NO:6) and the sugar biosynthesis enzymes encoded by the *ascF* gene of *Yersinia pseudotuberculosis* [Thorson *et al.*, *J. Bacteriol.*, 176:5483 (1994)], (SEQ ID NO:13), the *rfbJ* gene of *Salmonella enterica* [Jiang *et al.*, *Mol. Microbiol.*, 5:695 (1991)]. (SEQ ID NO:14), the *strL* gene of *Streptomyces griseus* [Pissowotzki *et al.*, *Mol. Gen. Genet.* 241:193 (1993)] (SEQ ID NO:15) and the galE gene of *Escherichia coli* [Lemaire and Hill, *Nucl. Acids Res.* 14:7705 (1986)] (SEQ ID NO:16). In this and all other Figures in which amino acid sequence identity is compared capitalized letters represent consensus (identical) amino acids between species or amino acids which are conservative substitutions for the consensus residues. Also in each Figure, the sequence identified as "consensus" is merely a convenient representation of conserved amino acids and is not intended as a representation of any existing polypeptide sequence.

FIG. 5B illustrates the amino acid sequence identity between the sugar biosynthesis enzyme encoded by the *eryBVII* gene of *Sac. erythraea* (SEQ ID NO:12) and the sugar biosynthesis enzymes encoded by the *strM* gene of *Streptomyces griseus* [Pissowotzki *et al.*, *Mol. Gen. Genet.* 241:193 (1993)] (SEQ ID NO:17), the *rfbC* gene of *Salmonella enterica* [Jiang *et al.*, *Mol. Microbiol.*, 5:695 (1991)] (SEQ ID NO:18), the *rfbF* gene of *Yersinia entercolitica* [Zhang *et al.*, *Mol. Microbiol.*, 9:309 (1993)] (SEQ ID NO:19), and the *ascE* gene of *Yersinia pseudotuberculosis* [Thorson *et al.*, *J. Bacteriol.*, 176:5483 (1994)] (SEQ ID NO:20).

FIG. 5C illustrates the amino acid sequence identity between the sugar biosynthesis enzyme encoded by the *eryCIV* gene of *Sac. erythraea* (SEQ ID NO:10) and the sugar biosynthesis enzymes encoded by the *eryCI* gene of *Sac. erythraea* [Dhillon *et al.*, *Mol. Microbiol.*, 3:1405 (1989)] (SEQ ID NO:21), the *ascC* gene of *Yersinia pseudotuberculosis* [Weigel *et al.*, *Biochemistry*, 31:2129 (1992), Thorson *et al.*, *J. Am. Chem. Soc.*, 115:6993 (1993), Thorson *et al.*, *J. Bacteriol.*, 176:5483 (1994)] (SEQ ID NO:22), the *dnrJ* gene of *Streptomyces peucetius* [Stutzman-Engwall *et al.*, *J. Bacteriol.*, 174:144 (1992)] (SEQ ID NO:23), the *prg1* gene of *Streptomyces alboniger* [Lacalle *et al.*, *EMBO J.*, 11:785 (1992)] (SEQ ID NO:24), and the *strS* gene of *Streptomyces griseus* [Distler *et al.*, *Gene*, 115:105 (1992)] (SEQ ID NO:25).

FIG. 5D illustrates the amino acid sequence identity between the sugar biosynthesis enzymes encoded by the *eryBV* and *eryCIII* genes of *Sac. erythraea* (SEQ ID NO:7 and SEQ ID NO:4 respectively) and the sugar biosynthesis enzyme encoded by the *dnrS* gene of *Streptomyces peucetius* [Otten *et al.*, *J. Bacteriol.*, 177:6688 (1995)] (SEQ ID NO:26).

FIG. 5E illustrates the amino acid sequence identity between the sugar biosynthesis enzyme encoded by the *eryCVI* gene of *Sac. erythraea* (SEQ ID NO:8) and the sugar biosynthesis enzymes encoded by the *srmX* gene of *Streptomyces ambofaciens* [Geistlich *et al.*, *Mol. Microbiol.*, 6:2019 (1992)] (SEQ ID NO:27), the *rdmD* gene of *Streptomyces purpurascens* [GenBank Accession: U10405] (SEQ ID NO:28) and the glycine methyltransferase of *Rattus norvegious* [Ogawa *et al.*, *Eur. J. Biochem.* 168:141 (1987)] (SEQ ID NO:29).

FIG. 6A through 6D illustrate the compounds conceivably formed in Examples 1-4 respectively and are representative of compounds formed from Type I (FIG 6A), Type II (FIG. 6B), and Type III (FIGS. 6C and 6D) alterations.

FIG. 7 illustrates the construction of the expression plasmid pASX2 described in Example 2. For FIGS 7-13 the following abbreviations have been used: *amp*, ampicillin resistance gene; *tsr*, thiostrepton resistance gene; ROP, repressor of plasmid synthesis gene; *eryBI*, *eryBII*, *eryBIII*, *eryBIV*, *eryBV*, *eryBVI*, *eryBVII*, *eryCI*, *eryCII*, *eryCIII*, *eryCIV*, *eryCV*, and *eryCVI*, the erythromycin biosynthetic genes involved in the synthesis of mycarose or its attachment to the macrolide ring (*eryB*) or the synthesis of desosamine or its attachment to the macrolide ring (*eryC*) [the thin arrows above a gene indicate its relative size and the direction of transcription]; ori-*E. coli*, an origin of DNA replication that functions in *E. coli*, in the specific examples the ColE1 origin; ori-*Streptomyces*, an origin of DNA replication that functions in *Streptomyces*, in the specific examples the pJV1 origin [Servin-Gonzalez *et al.*, *Microbiology*, 141:2499 (1995)]; p-*ermE**, a modified promoter for the erythromycin resistance gene; t-fd, the gene VIII transcription terminator of bacteriophage fd; PCR, polymerase chain reaction. Restriction enzyme sites have been indicated by their standard commercial names (i.e. *Bam*HI, *Eco*RI, etc). The abbreviations appended to the large arrows in the plasmid synthetic schemes summarize each of the steps involved the plasmid constructions. These steps are described fully in the relevant Examples.

FIG. 8 illustrates the construction of the *eryBVII* antisense expression plasmid pASBVII described in Example 2.

FIG. 9A illustrates the construction of the carrier plasmid pK1.

FIG. 9B-E illustrates the construction of plasmid pKB6 which carries all of the *eryB* genes and is described in Example 3.

FIG. 10 illustrates the construction of expression plasmid pX1 described in Example 3.

FIG. 11 illustrates the construction of the *eryB* expression plasmids pXSB6 and pXB6 described in Example 3.

FIG. 12A-B illustrate the construction of plasmid pKC4 which carries all of the *eryC* genes described in Example 4.

FIG. 13 illustrates the construction of the *eryC* expression plasmids pXSC4 and pXC4 described in Example 4.

Detailed Description of the Invention

I. The Invention

The present invention provides isolated and purified polynucleotides that encode enzymes or fragments thereof responsible for the biosynthesis of polyketide-associated sugars or their attachment to polyketides, vectors containing those polynucleotides, host cells transformed with those vectors, a process of making novel glycosylated polyketides using those polynucleotides and vectors, and isolated and purified recombinant polypeptides and polypeptide fragments thereof.

II. Definitions

For the purposes of the present invention as disclosed and claimed herein, the following terms are defined.

The term "polyketide" as used herein refers to a large and diverse class of natural products, including but not limited to antibiotic, antifungal, anticancer, and anti-helminthic compounds. Antibiotics include, but are not limited to anthracyclines and macrolides of different types (polyenes and avermectins as well as classical macrolides such as erythromycins).

The term "glycosylated polyketide" refers to any polyketide that contains one or more sugar residues.

The term "glycosylation-modified polyketide" refers to a polyketide having a changed glycosylation pattern or configuration relative to that particular polyketide's unmodified or native state.

The term "polyketide-producing microorganism" as used herein includes any microorganism that can produce a polyketide naturally or after being suitably engineered (i.e. genetically). Examples of actinomycetes and the polyketides they naturally produce include but are not limited to those listed in Table 1 below (see Hopwood, D.A. and Sherman, D.H., *Annu. Rev. Genet.*, 24:37-66 (1990) incorporated herein by reference).

Table 1

Organism	Polyketide Produced
Saccharopolyspora erythraea	Erythromycin
Streptomyces ambofaciens	Spiramycin
Streptomyces avermitilis	Avermectin
Streptomyces fradiae	Tylosin
Streptomyces griseus	Candicidin, monactin, griseusin
Streptomyces violaceoniger	Granaticin
Streptomyces thermotolerans	Carbomycin
Streptomyces rimosus	Oxytetracycline
Streptomyces peucetius	Daunorubicin
Streptomyces coelicolor	Actinorhodin
Streptomyces glaucescens	Tetracenomycin
Streptomyces roseofulvus	Frenolicin
Streptomyces cinnamonensis	Monensin
Streptomyces curacoi	Curamycin
Amycolatopsis mediterranei	Rifamycin

Other examples of polyketide-producing microorganisms that produce polyketides naturally include various *Actinomadura* , *Dactylosporangium* and *Nocardia* strains.

The term "sugar biosynthesis genes" as used herein refers to sequences of DNA from *Saccharopolyspora erythraea* that encode sugar biosynthesis enzymes and is intended to include sequences of DNA from other polyketide-producing microorganisms which are identical or analogous to those obtained from *Saccharopolyspora erythraea.*

The term "sugar biosynthesis enzymes" as used herein refers to polypeptides which are involved in the biosynthesis and/or attachment of polyketide-associated sugars and their derivatives and intermediates.

The term "polyketide-associated sugar" refers to a sugar that is known to attach to polyketides or that can be attached to polyketides by the processes described herein.

The term "sugar derivative" refers to a sugar which is naturally associated with a polyketide but which is altered relative to the unmodified or native state; examples only include N-3-α-desdimethyl D-desosamine, D-mycarose, 4-keto-L-mycarose, 4-keto-D-mycarose, 3-desmethyl L-mycarose and 3-desmethyl D-mycarose.

The term "sugar intermediate" refers to an intermediate compound produced in a sugar biosynthesis pathway.

The term "*eryB*" as used herein refers to sequences of DNA that encode enzymes involved specifically in the biosynthesis of the deoxysugar L-mycarose.

The term "*eryC*" as used herein refers to sequences of DNA that encode enzymes involved specifically in the biosynthesis of the deoxysugar D-desosamine.

III. Polynucleotides

The organization of the segment of the *Saccharopolyspora erythraea* (*Sac. erythraea*) chromosome that determines the biosynthesis of erythromycin and the corresponding genes that determine the biosynthesis of the sugars L-mycarose and D-desosamine, designated *eryB* and *eryC*, respectively, are shown in FIG. 1A. It is seen that several genes are required for the biosynthesis of each of the sugars and that these genes are interspersed among one another. It is predicted that each gene encodes an enzyme that catalyzes one or a few steps in the biosynthesis of L-mycarose or D-desosamine from thymidine diphospho-4-keto-6 deoxyglucose (TDP-glucose); these steps are outlined in FIG. 2 and FIG. 3. In the case of L-mycarose, (shown in FIG. 2), these steps include: (1) C-2" deoxygenation , (2) C-2"/C-3" enoyl reduction, (3) C-5" epimerization, (4) C-3" C-methylation, (5) C-4" keto reduction, and (6) transfer to erythronolide B. For D-desosamine, shown in FIG. 3, these steps comprise (1) C-4'/3' isomerization, (2, 3) C-3' deoxygenation and reduction, (4) C-3' amination, (5, 6) N-3a' N-dimethylation, and transfer to mycarosyl erythronolide B.

This classification of genes (as belonging to either the *eryB* class or *eryC* class) was determined by first altering the wild type genes of interest in an erythromycin producing strain (i.e. *in vivo*) to inactivate their expression. The erythromycin products resulting from such alterations were then analyzed. Genes whose alterations caused an accumulation of erythronolide B (indicating a lack of L-mycarose, or failure to attach L-mycarose to the erythronolide ring) were classified as *eryB* genes; genes whose alterations caused an accumulation of 3-α-L-mycarosyl erythronolide B (indicating a lack of D-desosamine, or failure to attach D-desosamine to the 3-α-L-mycarosyl erythronolide B ring) were classified as *eryC* genes. Accordingly, it should be noted that all such genes identified herein as *eryB* or *eryC* are involved in the synthesis of L-mycarose or D-desosamine. The predicted

functional activities of the polypeptides encoded by *eryB* and *eryC* will be discussed in further detail below.

In one aspect then, the present invention provides isolated and purified *eryB* and *eryC* polynucleotides from *Sac. erythraea* that encode enzymes involved in the production of glycosylated polyketides. A polynucleotide of the present invention that encodes a sugar biosynthesis enzyme is an isolated single or double stranded polynucleotide having a nucleotide sequence which comprises (a) a nucleotide sequence selected from the group consisting of (i) the sense sequence of FIG. 4A (SEQ ID NO:1) from about nucleotide position 54 to about nucleotide position 1136; (ii) the sense sequence of SEQ ID NO:1 from about nucleotide position 1147 to about nucleotide position 2412; (iii) the sense sequence of SEQ ID NO:1 from about nucleotide position 2409 to about nucleotide position 3410 ; (iv) the sense sequence of FIG. 4B (SEQ ID NO:2) from about nucleotide position 80 to about nucleotide position 1048; (v) the sense sequence of SEQ ID NO:2 from about nucleotide position 1048 to about nucleotide position 2295; (vi) the sense sequence of SEQ ID NO:2 from about nucleotide position 2348 to about nucleotide position 3061; (vii) the sense sequence of SEQ ID NO:2 from about nucleotide position 3214 to about nucleotide position 4677; (viii) the sense sequence of SEQ ID NO:2 from about nucleotide position 4674 to about nucleotide position 5879; (ix) the sense sequence of SEQ ID NO:2 from about nucleotide position 5917 to about nucleotide position 7386; and (x) the sense sequence of SEQ ID NO:2 from about nucleotide position 7415 to about nucleotide position 7996;

(b) sequences complementary to the sequences of (a),

(c) sequences that, when expressed, encode polypeptides encoded by the sequences of (a), and

(d) analogous sequences that hybridize under stringent conditions to the sequences of (a).

A preferred polynucleotide is a DNA molecule. In another embodiment, the polynucleotide is an RNA molecule.

The nucleotide sequence and deduced amino acid residue sequences of the sugar biosynthesis genes are set forth in FIG. 4A(1-4) and FIG. 4B(1-9). The nucleotide sequences of FIG. 4A(1-4) (SEQ ID NO:1) and FIG. 4B(1-9) (SEQ ID NO:2) represent full length DNA clones of the sense strand of two distinct clusters of sugar biosynthesis genes and are intended to represent both the sense strand (shown on top) and its complement. The amino acid sequences depicted below the sense strand correspond to polypeptides encoded by a nucleotide sequence selected from the group consisting of (i) the sense strand of SEQ ID NO:1 from about nucleotide position 54 to about nucleotide position 1136 (ii) the sense sequence of SEQ ID NO:1 from about nucleotide position 1147 to about nucleotide position 2412, (iii) the sense sequence of SEQ ID NO:1 from about nucleotide position 2409 to about

nucleotide position 3410, (iv) the sense sequence of SEQ ID NO:2 from about nucleotide position 80 to about nucleotide position 1048, (v) the sense sequence of SEQ ID NO:2 from about nucleotide position 1048 to about nucleotide position 2295, (vi) the sense sequence of SEQ ID NO:2 from about nucleotide position 2348 to about nucleotide position 3061, (vii) the sense sequence of SEQ ID NO:2 from about nucleotide position 3214 to about nucleotide position 4677, (ix) the sense sequence of SEQ ID NO:2 from about nucleotide position 5917 to about nucleotide position 7386 and (x) the sense sequence of SEQ ID NO:2 from about nucleotide position 7415 to about nucleotide position 7996. The polypeptides encoded by the nucleotide sequences of (i)-(x) above are set forth as SEQ ID NO:3-SEQ ID NO:12 respectively.

The present invention also contemplates analogous DNA sequences which hybridize under stringent hybridization conditions to the DNA sequences set forth above. Stringent hybridization conditions are well known in the art and define a degree of sequence identity greater than about 80%-90%. The modifier "analogous" refers to those nucleotide sequences that encode analogous polypeptides (i.e. in relation to a sugar biosynthesis enzyme), analogous polypeptides being those which have only conservative differences and which retain the conventional characteristics and activities of sugar biosynthesis enzymes. (A more detailed description of analogous polypeptides is provided below). The present invention also contemplates naturally occurring allelic variations and mutations of the DNA sequences set forth above so long as those variations and mutations code, on expression, for a sugar biosynthesis gene of this invention as set forth hereinafter.

As is well known in the art, because of the degeneracy of the genetic code, there are numerous other DNA and RNA molecules that can code for the same polypeptides as those encoded by the aforementioned sugar biosynthesis genes and fragments thereof. The present invention, therefore, contemplates those other DNA and RNA molecules which, on expression, encode the polypeptides of SEQ ID NO:3-SEQ ID NO:11 or fragments thereof. Having identified the amino acid residue sequence encoded by a sugar biosynthesis gene, and with knowledge of all triplet codons for each particular amino acid residue, it is possible to describe all such encoding RNA and DNA sequences. DNA and RNA molecules other than those specifically disclosed herein and, which molecules are characterized simply by a change in a codon for a particular amino acid, are within the scope of this invention.

The 20 common amino acids and their representative abbreviations, symbols and codons are well known in the art (see for example, Molecular Biology of the Cell, Second Edition, B. Alberts *et al.*, Garland Publishing Inc., New York and London, 1989). As is also well known in the art, codons constitute triplet sequences of nucleotides in mRNA molecules and as such, are characterized by the base uracil (U) in place of base thymidine (T) which is present in DNA molecules. A simple change in a codon for the same amino acid residue

within a polynucleotide will not change the structure of the encoded polypeptide. By way of example, it can be seen from SEQ ID NO:1 that an AGC codon for serine exists at nucleotide positions 126-128 and again at positions 420-422 and 561-563. However, it can also be seen from that same sequence that serine can be encoded by a TCG codon (see eg. nucleotide positions 192-194) and a TCC codon (see e.g., nucleotide positions 204-206). Substitution of the latter codons for serine with the AGC codon for serine, or *visa versa*, does not substantially alter the DNA sequence of SEQ ID NO:1 and results in production of the same polypeptide. In a similar manner, substitutions of the recited codons with other equivalent codons can be made in a like manner without departing from the scope of the present invention.

A polynucleotide of the present invention can also be an RNA molecule. An RNA molecule contemplated by the present invention is complementary to or hybridizes under stringent conditions to any of the DNA sequences set forth above. Exemplary and preferred RNA molecules are mRNA molecules that encode sugar biosynthesis enzymes of this invention.

IV. Polypeptides

In another aspect, the present invention provides polypeptides which are reasonably believed to be sugar biosynthesis enzymes. A sugar biosynthesis enzyme of the present invention is a polypeptide of about 21 kdal to about 47 kdal. As set forth in FIG. 5A-5E, analogs of the predicted polypeptides encoded by certain *eryB* and *eryC* genes have been identified in various species and their sequences compared using the PRETTY routine (Genetics Computer Group (GCG) Sequence Analysis Software Package, Madison, WI). Due to the degree of amino acid sequence identity existing between the polypeptides of these other sugar biosynthesis genes and the polypeptides encoded by the *eryB* and *eryC* genes, certain enzymatic activities can reasonably be attributed to the *eryB* and *eryC* polypeptides.

By way of example, analogs of the polypeptide encoded by the *eryBIV* gene have been identified in *Yersinia pseudotuberculosis*,*Salmonella enterica*, *Streptomyces griseus* and *Escherichia coli* (see FIG. 5A). The various analogs have been identified with from 290-328 amino acid residues and are characterized by a low degree of amino acid sequence identity. (For example, the identity between the sugar biosynthesis enzyme encoded by the *eryBIV* gene of *Sac. erythraea* and the sugar biosynthesis enzyme encoded by the *galE* gene of *E. coli* is 20% at the amino acid level). However, a conserved amino acid sequence motif, G x x G x x G (where G represents the amino acid glycine and x represents any other amino acid residue) is found within the first 30 amino acid residues of all analogs shown. Since the polypeptide encoded by the *galE* gene has been shown to be an epimerase (whose mechanism includes a ketoreduction (Bauer *et al.*, *Proteins* 12:372 (1992)), the *eryBIV* gene product is

reasonably predicted to be a ketoreductase.

As set forth in FIG. 5B analogs of the sugar biosynthesis enzyme encoded by the *eryBVII* gene have been identified in *Streptomyces griseus Salmonella enterica, Yersinia entercolitica* and *Yersinia pseudotuberculosis*. The various analogs have been identified with from 183-200 amino acid residues and are characterized by a moderate degree of amino acid identity. By way of example, the identity at the amino acid level between the sugar biosynthesis enzyme encoded by the *eryBVII* gene of *Sac. erythraea* and the sugar biosynthesis enzyme encoded by the *rfbC* gene of *Salmonella enterica* or the *strM* gene of *Streptomyces griseus* is 37% and 61%, respectively. Furthermore, a common characteristic of these particular polypeptides (including that of *eryBVII*), is that they are only associated with L-sugar biosynthesis and not with D-sugar biosynthesis. Thus the gene product of *eryBVII* is reasonably predicted to function as a C-5 epimerase which converts the stereochemistry of the sugar from the "D" configuration to the "L" configuration.

As set forth in FIG. 5C analogs of the sugar biosynthesis enzyme encoded by the *eryCIV* gene have been identified in *Sac. erythraea* and *Yersinia pseudotuberculosis*. As set forth in FIG. 5C, the predicted amino acid sequences of the protein products of *eryCI* and *eryCIV* share 34% sequence identity to each other, 27% and 25% respectively to the predicted amino acid sequence encoded by *ascC* from *Yersinia pseudotuberculosis*. The enzyme encoded by *ascC* has been shown to remove a hydroxyl group located at the C-3 position of L-ascarylose (Liu and Thorson, *Annu. Rev. Microbiol.* 48:223 (1994)). Thus, at least one of the polypeptides encoded by *eryCI* or *eryCIV* is predicted to be an enzyme which functions in deoxygenation reactions.

Furthermore, the enzyme encoded by the *ascC* gene requires the biochemical cofactor pyridoxamine, which is the same cofactor used in biochemical transamination reactions. Consequently, it has been proposed that some protein analogs (such as dnrJ from *Streptomyces peucetius*, prg1 from *Streptomyces alboniger* and strs from *Streptomyces griseus*) having a moderate degree of sequence similarity to the polypeptide encoded by *ascC* function as transaminases in amino sugar biosynthesis (Thorson *et al., J. Am. Chem. Soc.* 115:6993 (1993)). Since the biosynthesis of D-desosamine requires both deoxygenation and transamination, it is reasonable to predict that at least one of the polypeptides encoded by the *eryCI* or *eryCIV* genes functions in transamination reactions.

As set forth in FIG. 5D the predicted polypeptides encoded by *eryBV* and *eryCIII* share 43% identity at the amino acid level and as such, may be assumed to have similar activities with respect to their particular sugars. However, as shown in FIGS. 2 and 3, there are no common steps in the proposed pathways of L-mycarose and D-desosamine biosynthesis. Rather than having similar sugar biosynthesis functions, these polypeptides are predicted to be nucleotidyl-sugar transferases which, (in *Sac. erythraea* at least), function to

attach L-mycarose and D-desosamine to erythronolide B and 3-α-mycarosylerythronolide B, respectively.

As set forth in FIG. 5E analogs of the polypeptide encoded by the *eryCVI* gene have been identified in *Streptomyces ambofaciens*, *Streptomyces purpurascens*, and *Rattus norvegicus*. The various analogs have been identified with from 237-293 amino acid residues and are characterized by a low to moderate degree of amino acid identity. By way of example, the identity between the polypeptide encoded by the *eryCVI* gene of *Sac. erythraea* and the glycine methyltransferase of *Rattus norvegicus* is 26% at the amino acid level. Furthermore these sugar biosynthesis enzymes share a common sequence motif, LDVACGTG (SEQ ID NO:30 = amino acid positions 64-71 in the consensus sequence in FIG. 5E), with rat glycine methyltransferase whose biochemical function is known (Ogawa *et al.*, *Eur. J. Biochem.* 168:141 (1987)). Thus these polypeptides are predicted to be N-methyltransferases.

In another aspect, the present invention provides a recombinant C-4" keto reductase from *Sac. erythraea* . A recombinant *Sac. erythraea* C-4" ketoreductase of the present invention is a polypeptide of about 322 or less amino acid residues. A preferred recombinant *Sac. erythraea* C-4" ketoreductase is that encoded by the nucleotide sequence of SEQ ID NO:2 from about nucleotide position 80 to about nucleotide position 1048.

The present invention also contemplates amino acid residue sequences that are substantially duplicative of the sequences set forth herein such that those sequences demonstrate like biological activity to disclosed sequences. Such contemplated sequences include those analogous sequences characterized by a minimal change in amino acid residue sequence or type (e.g., conservatively substituted sequences) which insubstantial change does not alter the fundamental nature and biological activity of the aforementioned sugar biosynthesis enzymes.

It is well known in the art that modifications and changes can be made in the structure of a polypeptide without substantially altering the biological function of that peptide. For example, certain amino acids can be substituted for other amino acids in a given polypeptide without any appreciable loss of function. In making such changes, substitutions of like amino acid residues can be made on the basis of relative similarity of side-chain substituents, for example, their size, charge, hydrophobicity, hydrophilicity, and the like.

As detailed in United States Patent No. 4,554,101, incorporated herein by reference, the following hydrophilicity values have been assigned to amino acid residues: Arg (+3.0); Lys (+3.0); Asp (+3.0); Glu (+3.0); Ser (+0.3); Asn (+0.2); Gln (+0.2); Gly (0); Pro (-0.5); Thr (-0.4); Ala (-0.5); His (-0.5); Cys (-1.0); Met (-1.3); Val (-1.5); Leu (-1.8); Ile (-1.8); Tyr (-2.3); Phe (-2.5); and Trp (-3.4). It is understood that an amino acid residue can be substituted for another having a similar hydrophilicity value (e.g., within a value of plus or

minus 2.0) and still obtain a biologically equivalent polypeptide.

In a similar manner, substitutions can be made on the basis of similarity in hydropathic index. Each amino acid residue has been assigned a hydropathic index on the basis of its hydrophobicity and charge characteristics. Those hydropathic index values are: Ile (+4.5); Val (+4.2); Leu (+3.8); Phe (+2.8); Cys (+2.5); Met (+1.9); Ala (+1.8); Gly (-0.4); Thr (-0.7); Ser (-0.8); Trp (-0.9); Tyr (-1.3); Pro (-1.6); His (-3.2); Glu (-3.5); Gln (-3.5); Asp (-3.5); Asn (-3.5); Lys (-3.9); and Arg (-4.5). In making a substitution based on the hydropathic index, a value of within plus or minus 2.0 is preferred.

V. Production of novel glycosylated polyketides

In another aspect, the present invention comprises a general procedure for producing novel polyketide structures *in vivo* by selectively altering, inactivating, or augmenting the genetic information of the organism that naturally produces a related polyketide. That is, in the present invention, novel polyketides of desired structure are produced by manipulation of the *eryB* and/or *eryC* genes followed by their introduction into various polyketide-producing microorganisms. These manipulations result in the formation of "glycosylation-modified" polyketides (i.e. polyketides having an altered glycosylation pattern or configuration relative to their native state). For example, "glycosylation-modified" polyketides are those which have additional sugar groups attached (where none previously existed), different sugars (such as sugar intermediates) attached in place of the natural sugars or lack sugar groups (at positions where sugar groups previously existed).

In the case of Type I and Type II alterations (further described below) glycosylation-modified polyketides may arise though mechanisms which cause either (1) the non-production of the sugar attachment enzyme (i.e. the enzyme involved in attachment of a sugar to the the polyketide structure) or (2) the non-production of a sugar biosynthesis enzyme. In the first instance, the sugar will not be attached to the polyketide since the enzyme which functions to attach the sugar will be lacking. In the second situation, a sugar intermediate from the biosynthesis pathway will be produced (depending on which enzyme is lacking) and attached to the polyketide provided it is recognized as a suitable substrate by the sugar attachment enzyme; alternatively, it will not be recognized and therefore, not attached. In the case of Type III alterations (also described in detail below), glycosylation-modified polyketides arise via attachment of additional or different sugars (i.e. not normally found in a particular polyketide-producing strain) to the polyketide. It should be noted, that these postulated mechanisms are simply provided to enhance understanding of the novel processes described herein; the actual mechanisms by which the Type I, II and III alterations produce glycosylation-modified polyketides is not presently known.

In the first type of alteration (referred to herein as Type I alterations), genetically

altered *ery*B and/or *ery*C genes are introduced into the chromosome of *Sac. erythraea* or another glycosylated polyketide-producing organism that also produces L-mycarose, D-desosamine, or their closely related derivatives such as mycaminose (4-hydroxy D-desosamine). The genetic alteration of an *eryB* and/or *eryC* gene is such that it causes a non-functional enzyme to be synthesized. Once introduced into an appropriate strain, the altered gene replaces its corresponding wild type gene causing the strain to lose the ability to produce a particular enzymatic activity involved in sugar biosynthesis. As a result, a glycosylation-modified polyketide is produced via either of the mechanisms previously described for a Type I alteration.

In a Type I change described herein, a specific mutation in an *ery*B and/or *ery*C gene of the *Sac. erythraea* chromosome is accomplished by a three step process which involves: 1) specifically altering the DNA sequence of a desired sugar biosynthesis gene, 2) subcloning the altered sequence into a suitable vector capable of recombining in the chromosome of an appropriate host and 3) introducing the vector containing the subcloned sequence into the appropriate host so that exchange of the wild type allele with the mutated one will occur. The first step is accomplished using standard recombinant DNA techniques to effect a deletion, base pair conversion or frame-shift in the DNA sequence. The second step, which also employs standard recombinant techniques, involves subcloning the altered sequence into a vector which does not replicate in *Sac. erythraea* or the desired host. In the final step, the vector is introduced into a suitable host, where by the process of gene replacement, the altered allele replaces the wild-type one. All techniques employed in a Type I change are well known to those of ordinary skill in the art.

Example 1 illustrates the process of gene replacement of an *eryB* gene. As Example 1 shows, the *eryB* gene of interest is mutated and along with adjacent upstream and downstream DNA sequences, cloned into a non-replicating *Sac. erythraea* plasmid vector. The vector carrying the mutated allele and adjoining DNA is then introduced into the host strain by the process of protoplast transformation. Transformants are regenerated under selective conditions (i.e. conditions that require expression of a particular plasmid marker) in order to induce recombination of the plasmid into the host cell chromosome. In other words, since the plasmid does not replicate autonomously, it must reside in the chromosome to be maintained in the cell and to express a particular marker under selective conditions. Insertion is achieved when the regenerated cells undergo a single homologous recombination between one of the two DNA segments that flank the mutation on the plasmid and its homologous counterpart in the chromosome. The cells are then grown without selection for the marker which induces plasmid loss from the chromosome. This loss arises after the cells have undergone a second recombination between the second DNA segment that flanks the mutation and its homologous chromosomal counterpart. This second recombinational event

results in the loss of the plasmid sequences and the wild type allele from the chromosome; the mutant allele however is retained.

In a variation of a Type I change, the non-production of the sugar biosynthesis enzyme (or attachment enzyme) may be achieved by the alternative mechanisms of promoter inactivation and/or transcriptional terminator insertion. These variations do not effect the gene sequence itself but rather regulatory mechanisms involved in gene transcription. "Promoter" as used herein refers to that region of a DNA molecule which controls the initiation of RNA transcription. Such regions are known to bind RNA polymerases (i.e. the enzymes involved in synthesizing RNA molecules). This form of Type I change (i.e. promoter inactivation) involves two steps of 1) identifying the promoter region of the desired gene and 2) rendering the promoter region inoperable by mutation. As in the replacement mechanism described above such mutations may be effected by creating deletions in the promoter sequence or by base pair conversion. In the case where the promoter controls transcription of a single gene, inactivation of the promoter will eliminate expression of that particular gene; of course, where the promoter controls expression of an entire operon (i.e. a series of genes whose expression is controlled by a single promoter), promoter inactivation will effectively eliminate expression of all genes in that operon.

In a similar manner, the non-production of a sugar biosynthesis enzyme (or attachment enzyme) may arise from inserting a transcriptional terminator upstream from the gene to be inactivated. A "transcriptional terminator" as used herein is a nucleotide sequence which signals RNA polymerase to cease transcription. An example of a transcriptional terminator is a palindromic sequence capable of forming a stem-loop structure that is followed by a stretch of U residues (for example the transcriptional terminator that follows gene VIII of bacteriophage fd (Beck and Zink, *Gene*, 16:35 (1981)). Effecting a change in production of a sugar biosynthesis gene by this process involves 1) identifying of the gene or genes of interest (in the case of an operon arrangement) to be inactivated and 2) cloning a transcriptional terminator sequence in a region of the DNA upstream from such gene(s). A transcriptional terminator will cause the polymerase involved in RNA transcription to stop (at or near the signaling region) thereby preventing transcription of any downstream sequences. Thus, changes such as promoter inactivation and transcriptional insertion, which directly effect expression of sugar biosynthesis genes are also intended to be within the scope of the invention.

In the second case (referred to herein as Type II alterations) *eryB* and/or *eryC* genes are arranged on a vector in an antisense orientation relative to a promoter capable of allowing expression of the gene in *Sac. erythraea* or *Streptomyces*. The vector is then introduced into a polyketide producing microorganism. As a result of this vector construction, antisense messenger RNA (mRNA) is produced which interferes with the translation of the wild-type

mRNA. Similarly to the Type I manipulation, novel glycosylation modified polyketides will be produced in which the normal mycarose, desosamine, and/or closely related sugar residue is lacking or is substituted by a sugar intermediate.

In a Type II change, inactivation of the *eryB* and/or *eryC* genes by antisense expression is accomplished by a two step procedure in which (1) a specific sugar biosynthesis gene is subcloned into an expression vector in an antisense (i.e. reverse) orientation; and (2) the anti-sense expression vector is introduced into the desired strain. The first step is accomplished using standard recombinant DNA techniques employing either *E. coli* or *Streptomyces* as the host, and an expression vector (capable of replicating in either host) that can be assembled to contain a *Streptomyces* promoter. *Streptomyces* promoters may be obtained from any commercially available *Streptomyces* plasmids or *Streptomyces- E. coli* shuttle plasmids. In step 2, the anti-sense expression vector is introduced into a suitable *Streptomyces* strain and the transformed cells are grown under selective conditions in order to maintain the expression palsmid in the cell.

As described in Example 2, the gene to be inactivated is subcloned in its reverse orientation downstream of a *Streptomyces* promoter (which is contained within a replicating *Sac. erythraea* plasmid). The plasmid carrying the antisense gene is then introduced into the host strain by protoplast transformation. Transformants are regenerated under selective conditions in order to maintain the autonomously replicating plasmid in the cells. Subsequent expression of the antisense gene causes the production of an antisense messenger RNA (mRNA) that is complementary to the mRNA of the native allele of the selected gene. Through standard nucleotide base pair interactions, the antisense mRNA and the native mRNA form an RNA duplex that occludes the ribosome binding site of the native mRNA. This interaction prevents ribosomal translation of the native mRNA and the corresponding synthesis of the enzyme encoded by that mRNA. In this way, specific enzymatic steps in sugar biosynthesis corresponding to the identity of the gene expressed in the antisense orientation are blocked leading to the production of novel sugar intermediates which, when attached to the polyketide ring of the host microorganism, give rise to novel glycosylation-modified polyketides. Alternatively, the antisense expression vector can be constructed using a non-replicating *Sac. erythraea* vector that includes flanking DNA from a nonessential region of the *Sac. erythraea* chromosome, such as the region immediately upstream from the *eryK* gene (FIG. 1). This vector can then be used to stably insert the antisense construction into the chromosome by homologous recombination in a fashion similar to that described for the construction of a Type I alteration.

In the third case (referred to herein as Type III alterations), novel glycosylation-modified polyketides of desired structure are produced by arranging all or a subset of the *eryB* and/or *eryC* genes on a replicating vector and introducing these genes *en bloc* into a

"distinct" polyketide-producing organism, ie. one other than the microorganism from which the *eryB* and/or *eryC* genes were taken. As an example, *eryB* and/or *eryC* genes may be taken from *Sac. erythreae* and introduced into *Streptomyces violaceoniger* or *Streptomyces venezuelae*. In this case, mycarose, desosamine, their biochemical intermediates and/or their closely related derivatives will be synthesized and attached at specific positions to polyketide compounds that do not necessarily carry these, or any, sugar residues. Some examples of novel glycosylated polyketides that may be produced in hosts that carry such manipulations are shown in FIG. 6.

In Type III changes, the genes for the biosynthesis of mycarose and/or desosamine are introduced into a polyketide-producing organism other than *Sac. erythraea* by another simple two step procedure: 1) all or a subset of the *eryB* and/or *eryC* genes are assembled together on a replicating plasmid downstream of a *Streptomyces* promoter; and 2) the plasmid is introduced into the polyketide-producing organism. Step 1 requires standard recombinant DNA manipulations employing *E. coli* and/or *Streptomyces* as the host. Step 2 requires one or more plasmids out of the several *Streptomyces* vectors or *E. coli-Streptomyces* shuttle vectors available, one or more promoters that function in *Streptomyces*, and a selection for the presence of the strain carrying the plasmid. As described in Examples 3 and 4, sets of the *eryB* and/or *eryC* genes are sequentially subcloned together on a replicating vector downstream of a suitable promoter that functions in the desired host. The plasmid carrying the grouped genes is then introduced into the host strain by electroporation or by transformation of protoplasts employing selection for a plasmid marker.

GENERAL METHODS

Materials, Plasmids, and Bacterial Strains

Restriction endonucleases, T4 DNA ligase, competent *E. coli* DH5α cells, X-gal, IPTG and plasmids pUC18, pUC19, and pBR322 were purchased from Bethesda Research Laboratories (BRL), Gaithersburg, MD. Vent$_R$® DNA polymerase was purchased from New England Biolabs (Beverly, MA). Plasmids pGEM®5Zf, pGEM®7Zf, and pGEM®11Zf were from Promega, Madison, WI, plasmids pIJ4070 and pIJ702 were obtained from the John Innes Institute, Norwich, England, and plasmids pWHM3 and pWHM4 (*J. Bacteriol.* 1989 171:5872) were obtained from C. R. Hutchinson, University of Wisconsin, Madison, WI. [α-^{32}P]dCTP, Hybond™-N nylon membranes, and Megaprime nick translation kits were from Amersham Corp., Chicago, IL. SeaKem® LE agarose and SeaPlaque® low gelling temperature agarose were from FMC Bioproducts, Rockland, ME. *E. coli* K12 strains carrying the *E. coli-Sac. erythraea* shuttle plasmids pWHM3 and pWHM4 (Vara *et al.*, *J*

Bacteriol, 171:5872 (1989)) and pAIX have been deposited at the Agricultural Research Culture Collection (NRRL) 1815 N. University Street, Peoria, Illinois 61604, as of December 5, 1995, under the terms of the Budapest Treaty and will be maintained for a period of thirty (30) years from the date of deposit, or for five (5) years after the last request for the deposit, or for the enforceable period of the U.S. patent, whichever is longer. Plasmids pWHM3, pWHM4 and pAIX were accorded the accession numbers NRRL B-21512, NRRL B-21513 and NRRL B-21514, respectively. *Sac. erythraea* strain NRRL2338 is also available from the Agricultural Research Service culture collection. *Staphylococcus aureus* Th^R (thiostrepton resistant) was obtained by plating 10^8 cells of *S. aureus* on agar medium containing 10 µg/ml thiostrepton and picking a survivor after 48 hr growth at 37°C. Thiostrepton was obtained from Sigma Chemical, St. Louis, MO. All other chemicals and reagents were from standard commercial sources unless otherwise specified.

DNA Manipulations

Standard conditions were employed for restriction endonuclease digestion, agarose gel-electrophoresis, isolation of DNA fragments from low melting agarose gels, DNA ligation, plasmid isolation from *E. coli* by alkaline lysis, and transformation of *E. coli* employing selection for ampicillin resistance (150 µg/ml) on LB agar plates (Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Plainview, NY, 1989). Total DNA from *Sac. erythraea* and *Streptomyces* species (including *S. fradiae, S. celestes, S. violaceoniger, S. hygroscopicus, S. venezuelae*) was prepared according to described procedures (Hopwood *et al.*, *Genetic Manipulation of Streptomyces, A Laboratory Manual*, John Innes Foundation, Norwich, UK (1985)). Transfer of DNA from agarose gels to Hybond™-N membranes and Southern analysis using Megaprime™ nick translated probes was performed according to the manufacturers instructions.

Amplification of DNA Fragments

Synthetic deoxyoligonucleotides were synthesized on an ABI Model 380A synthesizer (Applied Biosystems, Foster City, CA) following the manufacturers recommendations. Amplification of DNA fragments was performed by the polymerase chain reaction (PCR) using a Perkin Elmer GeneAmp® PCR System 9600. Reactions contained 100 pmol of each primer, 1 µg of template DNA (chromosomal DNA from *Sac. erythraea* NRRL2338), 2 units $Vent_R$® DNA polymerase in 100 µl volume of PCR buffer (10 mM KCl, 10 mM $(NH_4)_2SO_4$, 20 mM Tris-HCl (pH 8.8, @ 25°C), 2.5 mM $MgSO_4$, 0.1% Triton® X-100) containing dATP (200 µM), dTTP (200 µM), dCTP (250 µM), and dGTP (250 µM). The reaction mixture was subjected to 30 cycles. Each cycle consisted of one period of 35 sec at 96°C and one period of 2 min at 72°C. The reaction products were visualized and

purified from low melting agarose. The PCR primers described in the examples were derived from the nucleotide sequence of the *eryB* and *eryC* genes of FIG. 4.

Transformation and Gene Replacement in *Sac. erythraea*

Protoplasts of *Sac. erythraea* strains were prepared and transformed with miniprep DNA isolated from *E. coli* according to published procedures (Yamamoto *et al.*, *J Antibiotics*, 39:1304 (1986)). Non-integrative transformants, in the case of pWHM4 derivatives, were selected by regenerating the protoplasts and overlaying with thiostrepton (final concentration 20 µg/ml) as described (Weber *et al.*, *Gene*, 68:173 (1988)). Integrative transformants, in the case of pWHM3 derivatives, were selected on thiostrepton-containing agar plates (15 µg/ml) as described by Weber *et al.*, *Gene*, 68:173 (1988). Loss of the Th^R phenotype was monitored after two rounds of non-selective growth in SGGP media (Yamamoto *et al.*, *J Antibiotics*, 39:1304 (1986)) followed by protoplasting and serial dilution on non-selective agar media. Regenerated protoplasts were replica plated on thiostrepton-containing media. Th^S (thiostrepton-sensitive) colonies arose at a frequency of 10^{-1}. Retention of the mutant allele was established by Southern hybridization of several Th^S colonies.

Fermentation

Sac. erythraea or *Streptomyces* cells are inoculated into 100 ml SCM medium (1.5% soluble starch, 2.0% Difco Soytone, 0.15% Yeast Extract, 0.01% $CaCl_2$) and allowed to grow for 3 to 6 days. The entire culture is then inoculated into 10 liters of fresh SCM medium. The fermenter is operated for a period of 4 to 7 days at 32°C maintaining constant aeration and pH at 7.0. After the fermentation is complete, the cells are removed by centrifugation at 4°C and the fermentation beer is kept cold until further use. When antibiotic selection to maintain a plasmid, such as pXC4 or pXB6, is required, thiostrepton (10µg/ml) is added to both the 100 ml starter culture and the 10-liter fermenter.

The invention will be better understood in connection with the following examples, which are intended as an illustration of and not a limitation upon the scope of the invention. Both below and throughout the specification, it is intended that citations to the literature be expressly incorporated by reference.

Example 1: Construction and characterization of *Sac. erythraea* ERBIV that produces 4"-deoxy-4"-oxo-erythromycin A

A. Construction of Plasmid pRBIV: A 4.3 kb *Pst*I-*Hin*dIII fragment, which included

the *eryBIV* gene, was isolated from the plasmid pAIX5 and subcloned into *Pst*I-*Hin*dIII digested pUC19 to generate plasmid pUCBIV. After transformation and isolation of the plasmid from *E. coli*, the identity of pUCBIV was confirmed by digestion with *Mun*I which released a fragment of 370 bp. Plasmid pUCBIV was then cut with the restriction enzyme *Nco*I, the restriction site filled in with Klenow enzyme, and the plasmid religated to generate plasmid pNCOBIV, (which now carried a frameshift mutation in the *eryBIV* gene). After transformation and isolation of the plasmid from *E. coli*, the identity of pNCOBIV was confirmed by digestion with *Nsi*I and *Hin*dIII which released a fragment of 1.59 kb. (The *Nsi*I site was formed by the fill-in and religation of the *Nco*I site.) Finally, plasmid pNCOBIV was digested with *Hin*dIII and *Sst*I and the 3.2 kb fragment carrying the altered *eryBIV* gene was isolated and ligated into *Hin*dIII and *Sst*I digested pWHM3 to generate plasmid pRBIV. After transformation and isolation of the plasmid from *E. coli*, the identity of pRBIV was confirmed by digestion with *Kpn*I which released fragments of 5.2 kb, 4.4 kb, and 0.72 kb.

B. Construction of *Sac. erythraea* ERBIV: *Sac. erythraea* protoplasts were transformed with plasmid pRBIV and integrative transformants selected as described in General Methods. Resolution of the integrants by nonselective growth as described in General Methods yielded *Sac. erythraea* ERBIV in which the wild type copy of the *eryBIV* gene was replaced with the inactive mutant copy. Gene replacement was confirmed by Southern analysis of *Nco*I digested *Sac. erythraea* DNA and *Nco*I-*Nsi*I digested *Sac. erythraea* DNA using the 1.58 kb *Nco*I-*Hin*dIII fragment isolated from plasmid pUCBIV (coordinates 681-2214, FIG. 4B) as a probe. Wild type *Sac. erythraea* and wild type resolvants display a hybridizing DNA fragment of 2.75 kb when digested with either *Nco*I or *Nco*I-*Nsi*I, whereas *Sac. erythraea* strain ERBIV is characterized by hybridization to either a 16 kb DNA fragment or a 2.75 kb DNA fragment when digested with *Nco*I or *Nco*I-*Nsi*I, respectively.

C. Isolation, purification, and properties of 4"-deoxy-4"-oxo-erythromycin A from *Sac. erythraea* ERBIV: *Sac. erythraea* strain ERBIV is fermented for 4 days in SCM media as described in General Methods. The fermentation broth of *Sac. erythraea* ERBIV is then cooled to 4°C and adjusted to pH 4.0 and extracted once with methylene chloride. The aqueous layer is readjusted to pH 9.0 and extracted twice with methylene chloride and the combined basic methylene chloride extracts are concentrated to a solid residue. This is digested in methanol and chromatographed over a column of Sephadex LH-20 in methanol. Fractions are tested for bioactivity against a sensitive organism, such as *Staphylococcus aureus* Th^R, and active fractions are combined. The combined fractions are concentrated and the residue is digested in 10 ml of the upper phase of a solvent system consisting of n-heptane, benzene, acetone, isopropanol, 0.05 M, pH 7.0 aqueous phosphate buffer

(5:10:3:2:5, v/v/v/v/v), and chromatographed on an Ito Coil Planet Centrifuge in the same system. Active fractions are combined, concentrated and partitioned between methylene chloride and dilute ammonium hydroxide (pH 9.0). The methylene chloride layer is separated and concentrated to yield the desired product as a white foam.

Example 2: Construction and characterization of *Sac. erythraea* ER720(pASBVII) that produces 3-α-D-mycarosyl-5-ß-D-desosaminoyl-12-hydroxy-erythronolide B

A. Construction of plasmid pASX2 (see FIG. 7): The 290 bp *Eco*RI-*Bam*HI segment carrying the *ermE** promoter is isolated from plasmid pIJ4070 and ligated into *Eco*RI-*Bam*HI digested pWHM4 DNA to form pASX1. After transformation and isolation of the plasmid from *E. coli,* the identity of pASX1 is confirmed by digestion with *Apa*LI which releases fragments of 3.9 kb, 2.5 kb, 1.2 kb, 0.5 kb, and 0.4 kb. Two oligonucleotides of the sequences: SEQ ID NO:31 (5'-GATCCAGCGTCTGCAGGCATGCTCTAGATACAATTA AAGGCTCCTTTTGGAGCCTTTTTTTTTGGAGATTTTCAACGT-3') and SEQ ID NO:32 (5'-AGCTACGTTGAAAATCTCCAAAAAAAAAGGCTCCAAAA GGAGCCTTTAATTGTATCTAGAGCATGCCTGCAGACGCTG-3'), corresponding to the (+) and (-) strands of the bacteriophage fd gene VIII transcription terminator (t-fd) (Beck *et al.* (1978) *Nucl. Acids Res.* 5:4495])and including restriction enzyme sites for the enzymes *Pst*I, *Sph*I, and *Xba*I, and overhanging ends compatible with *Bam*HI and *Hin*dIII are synthesized and approximately 250 ng of each oligonucleotide are then mixed together in TE buffer and heated to 99°C for 1 min. The solution is cooled slowly to room temperature allowing the oligonucleotides to anneal due to self complementarity, and the annealed oligonucleotides are then ligated into *Bam*HI-*Hin*dIII digested pASX1 to give pASX2. After transformation and isolation of the plasmid from *E. coli,* the identity of pASX2 is confirmed by DNA sequencing of the 1.2 kb *Eco*RI-*Sal*I fragment that contains the *ErmE** promoter and the bacteriophage fd terminator.

B. Construction of plasmid pASBVII (see FIG. 8): The 598 base pair DNA segment that carries the *eryBVII* gene, comprising coordinates 7398-7996 (FIG. 4B), is amplified by PCR employing two oligonucleotides, SEQ ID NO:33 (5'-GATCGCATGCTCTAGAGTACG-TGAGCTGGCGGTGGCGGGC-3') and SEQ ID NO:34 (5'-GATCCGGATCCGCATGCTT-CACCTGCCGGTGCTGGCGGG-3'). After digestion of the purified PCR product with *Bam*HI-*Xba*I the PCR fragment was ligated to *Bam*HI-*Xba*I digested pASX2 to give pASBVII. After transformation and isolation of the plasmid from *E. coli,* the identity of pASBVII is verified by DNA sequencing of the 880 bp *Eco*RI-*Xba*I insert.

C. Construction of *Sac. erythraea* ER720(pASBVII): *Sac. erythraea* strain ER720

protoplasts are transformed with plasmid pASBVII and transformants are selected for with thiostrepton (15 μg/ml). To confirm transformation, total DNA is isolated from Th^R colonies and used to transform *E. coli*. After transformation and isolation of the plasmid from *E. coli*, the identity of pASBVII is verified by restriction analysis with the enzymes *Pvu*II and *Bam*HI which releases a 1.48 kb fragment. Those *Sac. erythraea* colonies that are found to contain pASBVII are designated *Sac. erythraea* ER720(pASBVII).

D. Isolation, purification, and properties of 3-α-D-mycarosyl-5-β-D-desosaminoyl-12-hydroxy-erythronolide B from *Sac. erythraea* ER720(pASBVII): *Sac. erythraea* ER720(pASBVII) is fermented for 3 days in SCM media with thiostrepton selection as described in General Methods. The fermentation broth is then cooled to 4°C and adjusted to pH 4.0 and extracted once with methylene chloride. The aqueous layer is readjusted to pH 9.0 and extracted twice with methylene chloride and the combined extracts are concentrated to a solid residue. This is digested in methanol and chromatographed over a column of Sephadex LH-20 in methanol. Fractions are tested for bioactivity against a sensitive organism, such as *Staphylococcus aureus* Th^R, and active fractions are combined. The combined fractions are concentrated and the residue is digested in 10 ml of the upper phase of a solvent system consisting of n-heptane, benzene, acetone, isopropanol, 0.05 M, pH 7.0 aqueous phosphate buffer (5:10:3:2:5, v/v/v/v/v), and chromatographed on an Ito Coil Planet Centrifuge in the same system. Active fractions are combined, concentrated and partitioned between methylene chloride and dilute ammonium hydroxide (pH 9.0). The methylene chloride layer is separated and concentrated to yield the desired product as a white foam.

Example 3: Construction and characterization of *Streptomyces antibioticus* ATCC 11891(pXB6) that produces 3-des-oleandrosyl-3-mycarosyl oleandomycin

A. Construction of plasmid pKB6 and intermediates (see FIG. 9)

i) Construction of plasmid pK1: The DNA sequences of pBR322 (GenBank Accession #: J01749) and pUC19 (GenBank Accession #: X02514) are known. The 805 nt DNA segment comprising coordinates 1673 through 2478 of pBR322 is amplified by PCR employing two oligodeoxynucleotides, SEQ ID NO:35 (5'-GATCACATGTTCTTTCCTG-CGTTATCCCCTG-3') and SEQ ID NO:36 (5'-GATCGGATCCATGCATGTCTAGAGCA-TCGCAGGATGCTGCTGGC-3'). After digestion of the purified PCR product with *Afl*III and *Bam*HI the fragment is ligated into *Afl*III and *Bam*HI digested pUC19 to give plasmid pK1. The identity of plasmid pK1, after transformation and isolation from *E. coli*, is verified by *Pvu*II digestion which releases fragments of 0.55 kb and 2.55 kb. Plasmid pK1 contains the ROP region of pBR322 that controls plasmid copy number.

ii) Construction of plasmid pKB1: The 2.24 kb DNA segment that carries the

eryBIV and *eryBV* genes, comprised between coordinates 56 and 2296 of the sequence presented in SEQ ID NO:2, is amplified by PCR employing two deoxyoligonucleotides, SEQ ID NO:37 (5'-GAATGCATCCTGGAAAGCGAGCAAATGCTCCGGTG-3') and SEQ ID NO:38 (5'-GATCTAGAGCTAGCCGGCGTGGCGGCGCGTG-3'). After digestion with *Nsi*I and *Xba*I the fragment is ligated into *Nsi*I and *Xba*I digested pK1 to yield plasmid pKB1, 5.3 kb in size. The identity of plasmid pKB1, after transformation and isolation from *E. coli*, is verified by *Kpn*I digestion which releases fragments of 0.72 kb, 1.14 kb and 3.42 kb.

iii) Construction of plasmid pKB2: The 1.56 kb DNA segment that carries the *eryBVI* gene, comprised between coordinates 3121 and 4677 of the sequence presented in SEQ ID NO:2, is amplified by PCR employing two deoxyoligonucleotides, SEQ ID NO:39 (5'-GATCGCTAGCCGTGACCGGACCCTTACAGTGAGTG-3') and SEQ ID NO:40 (5'-GATCTAGACTTAAGTCATCCGGCGGTCCTGGTGTAGACGGC-3'). After digestion with *Nhe*I and *Xba*I the fragment is ligated into *Nhe*I and *Xba*I digested pKB1 to give plasmid pKB2, 6.9 kb in size. The identity of plasmid pKB2, after transformation and isolation from *E. coli*, is confirmed by *Bam*HI digestion which releases fragments of 0.22 kb, 0.40 kb, 2.6 kb and 3.7 kb.

iv) Construction of plasmid pKB3: The 0.6 kb DNA segment that carries the *ery BVII* gene, comprised between coordinates 7385 and 7987 of the sequence presented in SEQ ID NO:2, is amplified by PCR employing two deoxyoligonucleotides, SEQ ID NO:41 (5'-GATCTTAAGAACCGGAGTTGCGAGTACGTGAGCTGGCG-3') and SEQ ID NO:42 (5'-GATCTAGACCTAGGTCACCTGCCGGTGCTGGCGGGCTC-3'). After digestion with *Afl*II and *Xba*I the fragment is ligated into *Afl*II and *Xba*I digested pKB2 giving plasmid pKB3, 7.5 kb in size. The identity of plasmid pKB3, after transformation and isolation from *E. coli*, is verified by *Pst*I digestion which releases fragments of 1.1 kb and 6.4 kb.

v) Construction of plasmid pKB4: The 1.0 kb DNA segment that carries the *eryBII* gene, comprised between coordinates 2385 and 3410 of the sequence presented in SEQ ID NO:1, is amplified by PCR employing two deoxyoligonucleotides, SEQ ID NO:43 (5'-GATCCTAGGCCGCAGGAAGGAGAGAACCACG-3') and SEQ ID NO:44 (5'-GATCTAGATTAATCACTGCAACCAGGCTTCCGGC-3'). Following digestion with *Avr*II and *Xba*I the fragment is ligated into *Avr*II and *Xba*I digested pKB3 yielding the desired plasmid pKB4. After transformation and isolation of the plasmid from *E. coli*, the identity of pKB4, 8.5 kb in size, is verified by *Bgl*II and *Eco*RI digestion which releases fragments of 0.41 kb, 1.6 kb, 3.1 kb and 3.4 kb.

vi) Construction of plasmid pKB5: The DNA sequence of *eryBIII* has been reported (Haydock *et al* (1991) *Mol Gen Genet* **230**:120). The 1.3 kb DNA segment that carries the *eryBIII* gene, comprised between coordinates 3965 and 5232 of the sequence depicted in Haydock *et al*, is amplified by PCR employing two deoxyoligonucleotides, SEQ

ID NO:45 (5'-GATTAATTGGCCGCGGCGCCGCGCTC-GTTATG-3') and SEQ ID NO:46 (5'-GATCTAGATAATTAATCATACGACTTCCAGTC-GGGGTAG-3'). After digestion with *Mse*I and *Xba*I the fragment is ligated into *Mse*I and *Xba*I digested pKB4 to give the desired plasmid pKB5, 9.8 kb in size. The identity of pKB5, after transformation and isolation from *E. coli*, is verified by *Pst*I digestion which releases fragments of 1.1 kb, 2.5 kb, and 6.1 kb, visualized by gel electrophoresis.

vii) Construction of plasmid pKB6: The *eryBI* gene has been mapped (Haydock *et al* (1991) *Mol Gen Genet* **230**:120) and the DNA sequence on both flanks of *eryBI* is known (Haydock *et al* (1991) *Mol Gen Genet* **230**:120) and GenBank Accession # M11200. The 2.5 kb DNA segment that carries the *eryBI* gene, comprised between coordinates 1.1 and 3.6 of the map presented in Haydock *et al.*, is amplified by PCR employing two deoxyoligonucleotides: SEQ ID NO:47 (5'-GATTAATTAATGATCA-AGCTGAAAATTGTTTGCATG-3') and SEQ ID NO:48 (5'-GATCTAGACTGCCGGCT-CAGCCTTCCCAGGTTCG-3'). After digestion with *Pac*I and *Xba*I the fragment is ligated into *Pac*I and *Xba*I digested pKB5 to give plasmid pKB6, 12.3 kb in size. The identity of pKB6, after transformation and isolation from *E. coli*, is verified by *Bam*HI digestion which releases fragments of 0.22 kb, 0.40 kb, 1.4 kb, 2.6 kb, 3.3 kb and 4.4 kb. Plasmid pKB6 carries all of the *eryB* genes, *eryBI-eryBVII*, that are involved in the biosynthesis of mycarose and its attachment to the polyketide.

B. Construction of Plasmid pXSB6 (see FIG. 11): The 9.2 kb *Nsi*I-*Xba*I segment of pKB6, prepared as described in Example 3(A)(vii) above, that carries all of the *eryB* genes is isolated and ligated into *Pst*I-*Xba*I digested pASX2, prepared as described in Example 2(A) above, to give plasmid pXSB6. After transformation and isolation of the plasmid from *E. coli*, the identity of pXSB6, 17.2 kb in size, is verified by the observation of fragments of 0.41 kb, 1.9 kb, and 14.9 kb after *Eco*RI digestion. Plasmid pXSB6 carries all of the *eryB* genes in a transcriptional fusion downstream of the *ermE** promoter on an *E. coli-Streptomyces* shuttle plasmid.

C. Construction of Plasmid pXB6

i) Construction of plasmid pN702 (see FIG. 10): Two oligonucleotides of the sequences: SEQ ID NO:49 5'-GGAATTCAGATCTATGCATTCTAGAA-3') and SEQ ID NO:50 (5'-CGCGTTCTAGAATGCATAGATCTGAATTCCTGCA-3') that include restriction enzyme sites for the enzymes *Eco*RI, *Bgl*II, *Nsi*I, and *Xba*I and overhanging ends compatible with *Pst*I and *Mlu*I are synthesized. Approximately 250 ng of each oligonucleotide are then mixed together in TE buffer and heated to 99°C for 1 min. After the solution is cooled slowly to room temperature allowing the oligonucleotides to anneal due to self complementarity, the annealed oligonucleotides are ligated into *Pst*I-*Mlu*I digested pIJ702 to yield the desired plasmid pN702. After transformation and isolation of the plasmid

from *Streptomyces lividans* 1326, the identity of plasmid pN702, 4.3 kb in size, is verified by the observation of fragments of 0.75 kb and 3.6 kb after *Eco*RI-*Bam*HI or *Xba*I-*Bam*HI digestion.

ii) Construction of plasmid pX1 (see FIG. 10): The 290 bp *Eco*RI-*Bam*HI segment that carries the *ermE** promoter is isolated from plasmid pIJ4070 and ligated into *Eco*RI-*Bgl*II digested pN702 to give plasmid pX1. The resulting mixture contains the desired plasmid pX1. After transformation and isolation of the plasmid from *Streptomyces lividans* 1326, the identity of plasmid pX1, 4.6 kb in size, is verified by the observation of fragments of 1.0 kb and 3.6 kb after *Nsi*I-*Bam*HI digestion.

iii) Construction of plasmid pXB6 (see FIG. 11): The 9.2 kb *Nsi*I-*Xba*I segment of pKB6, prepared as described in Example 3(A)(vii) above, that carries all of the *eryB* genes is isolated and ligated into *Nsi*I-*Xba*I digested pX1 to give the desired plasmid pXB6. After transformation and isolation of the plasmid from *Streptomyces lividans* 1326, the identity of plasmid pXB6, 13.8 kb in size, is verified by the observation of fragments of 0.41 kb, 1.9 kb, and 11.5 kb after *Eco*RI digestion. Plasmid pXB6 carries all of the *eryB* genes in a transcriptional fusion to the *ermE** promoter on a *Streptomyces* plasmid.

D. Construction of *Streptomyces antibioticus* ATCC 11891(pXB6): Approximately 500 μg of plasmid pXB6, isolated from *Streptomyces lividans* 1326(pXB6), are electroporated into the oleandomycin producer *Streptomyces antibioticus* ATCC 11891 and several of the resulting Thio^R colonies that appear on the R3M-agar plates containing thiostrepton are analyzed for their plasmid content. The presence of plasmid pXB6, 13.8 kb in size, is verified by the observation of fragments of 0.41 kb, 1.9 kb, and 11.5 kb after *Eco*RI digestion.

E. Isolation, purification, and properties of 3-des-oleandrosyl-3-mycarosyl oleandomycin from *Streptomyces antibioticus* ATCC 11891(pXB6): *Streptomyces antibioticus* ATCC 11891(pXB6) is fermented for 5 days in SCM media with thiostrepton selection as described in General Methods. The fermentation broth is then cooled to 4°C and adjusted to pH 4.0 and extracted once with methylene chloride. The aqueous layer is readjusted to pH 9.0 and extracted twice with methylene chloride and the combined extracts are concentrated to a solid residue. This is digested in methanol and chromatographed over a column of Sephadex LH-20 in methanol. Fractions are tested for bioactivity against a sensitive organism, such as *Staphylococcus aureus* Th^R, and active fractions are combined. The combined fractions are concentrated and the residue is digested in 10 ml of the upper phase of a solvent system consisting of n-heptane, benzene, acetone, isopropanol, 0.05 M, pH 7.0 aqueous phosphate buffer (5:10:3:2:5, v/v/v/v/v), and chromatographed on an Ito Coil Planet Centrifuge in the same system. Closely eluting active fractions are combined, concentrated and partitioned between methylene chloride and dilute ammonium hydroxide

(pH 9.0). The methylene chloride layer is separated and concentrated to yield the desired product as a white foam.

Example 4: Construction and characterization of *Streptomyces violaceoniger* NRRL 2834(pXC4) that produces 5-des-chalcosyl-5-desosaminoyl lankamycin

A. Construction of plasmid pKC4 and intermediates (see FIG. 12)

i) Construction of plasmid pKC1: The 2.4 kb DNA segment that carries the *eryCII* and *eryCIII* genes, comprised between coordinates 33 and 2413 of the sequence presented in SEQ ID NO:1, is amplified by PCR employing two deoxyoligonucleotides, SEQ ID NO:51 (5'-GAATGCATCTGGCTGGGCGGAGGGAATTCATG-3') and SEQ ID NO:52 (5'-GATCTAGACTTAAGTCATCGTGGTTCTCTCCTTCCTGC GGC-3'). After digestion with *Nsi*I and *Xba*I the purified PCR fragment is ligated into *Nsi*I and *Xba*I digested pK1 to give plasmid pKC1, 5.5 kb in size. The identity of plasmid pKC1, after transformation and isolation from *E. coli*, is verified by *Eco*RI digestion which releases fragments of 2.2 kb and 3.3 kb.

ii) Construction of plasmid pKC2: The 732 bp DNA segment that carries the *eryCVI* gene, comprised between coordinates 2331 and 3063 of the sequence presented in SEQ ID NO:2, is amplified by PCR employing two deoxyoligonucleotides, SEQ ID NO:53 (5'-GATCCTTAAGCTCCGGAGGGAGCAGGGATG-3') and SEQ ID NO:54 (5'-GATCTAGACCTAGGTCATCCGCGCACACCGACGAAC-3'). After digestion with *Afl*II and *Xba*I the purified PCR fragment is ligated into *Afl*II and *Xba*I digested pKC1 to give plasmid pKC2, 6.2 kb in size. The identity of plasmid pKC2, after transformation and isolation from *E. coli*, is verified by *Xba*I-*Eco*RI digestion which releases fragments of 0.95 kb, 2.2 kb and 3.1 kb.

iii) Construction of plasmid pKC3: The 2.7 kb DNA segment that carries the *eryCIV* and *eryCV* genes, comprised between coordinates 4650 and 7386 of the sequence presented in SEQ ID NO:2, is amplified by PCR employing two deoxyoligonucleotides, SEQ ID NO:55 (5'-GATCCTAGGCCGTCTACACCAGGACCGCCGG-3') and SEQ ID NO:56 (5'-GATCTAGATTAATCACCTTCCGCGCAGGAAGCCGC-3'). After digestion with *Avr*II and *Xba*I the purified PCR fragment is ligated into *Avr*II and *Xba*I digested pKC2 to yield plasmid pKC3, 9.0 kb in size. The identity of plasmid pKC3, after transformation and isolation from *E. coli*, is verified by *Sph*I digestion which releases fragments of 4.0 kb and 5.0 kb.

iv) Construction of plasmid pKC4: The DNA sequence of the *eryCI* gene has been determined (GenBank Accession #X15541). The 1.1 kb DNA segment that carries the *eryCI* gene, comprised between coordinates 38 and 1161 of the sequence indicated above, is

amplified by PCR employing two deoxyoligonucleotides, SEQ ID NO:57 (5'-GATCTTAAG-CCGCCACTCGAACGGACACTCG-3') and SEQ ID NO:58 (5'-GATCTAGATCAAGCCC-CAGCCTTGAGGG-3'). After digestion with *Mse*I and *Xba*I the fragment is ligated into *Mse*I and *Xba*I digested pKC3 to give plasmid pKC4, 10.1 kb in size. The identity of plasmid pKC4, after transformation and isolation from *E. coli*, is verified by *Kpn*I digestion which releases fragments of 0.15 kb, 0.31 kb, 4.1 kb and 5.5 kb. Plasmid pKC4 carries all of the *eryC* genes, *eryCI-eryCVI*, that are involved in the biosynthesis of desosamine and its attachment to the polydetide.

B. Construction of Plasmid pXSC4 (see FIG. 13): The 6.9 kb *Nsi*I-*Xba*I segment of pKC4 that carries all of the *eryC* genes is isolated and ligated into *Pst*I-*Xba*I digested pASX2, prepared as described in Example 2(A), to give the desired plasmid pXSC4, 14.9 kb in size, wherein all of the *eryC* genes are transcriptionally linked downstream of the *ermE** promoter on an *E. coli-Streptomyces* shuttle plasmid. The identity of plasmid pXSC4, after transformation and isolation from *E. coli*, is verified by the observation of fragments of 0.29 kb, 2.2 kb, and 12.4 kb after *Eco*RI digestion .

C. Construction of Plasmid pXC4 (see FIG. 13): The 6.9 kb *Nsi*I-*Xba*I segment of pKC4 that carries all of the *eryC* genes is isolated and ligated into *Nsi*I-*Xba*I digested pX1, prepared as described in Example 3(C)(ii), to give the desired plasmid pXC4, 11.5 kb in size, wherein all of the *eryC* genes are transcriptionally linked downstream of the *ermE** promoter on a *Streptomyces* plasmid. After transformation and isolation of the plasmid from *Streptomyces lividans* 1326, the identity of plasmid pXC4 is verified by the observation of fragments of 0.29 kb, 2.2 kb, and 9.0 kb after *Eco*RI digestion.

D. Construction of *Streptomyces violaceoniger* NRRL 2834(pXC4): Approximately 500 µg of the plasmid pXC4, isolated from *Streptomyces lividans* 1326(pXC4) , are electroporated into the lankamycin producer *Streptomyces violaceoniger* NRRL 2834 and several of the resulting Thio^R colonies that appear on the R3M-agar plates containing thiostrepton are analyzed for their plasmid content. The presence of plasmid pXC4 is verified by the observation of fragments of 0.29 kb, 2.2 kb, and 9.1 kb in size after *Eco*RI digestion of the plasmid.

E. Isolation, purification, and properties of 5-des-chalcosyl-5-desosaminoyl lankamycin: *S. violaceoniger* NRRL 2834(pXC4) is fermented for 5 days in SCM media with thiostrepton selection as described in General Methods. The fermentation broth is then cooled to 4°C and adjusted to pH 4.0 and extracted once with methylene chloride. The aqueous layer is readjusted to pH 9.0 and extracted twice with methylene chloride and the combined extracts are concentrated to a solid residue. This is digested in methanol and chromatographed over a column of Sephadex LH-20 in methanol. Fractions are tested for bioactivity against a sensitive organism, such as *Staphylococcus aureus* Th^R, and active

fractions are combined. The combined fractions are concentrated and the residue is digested in 10 ml of the upper phase of a solvent system consisting of n-heptane, benzene, acetone, isopropanol, 0.05 M, pH 7.0 aqueous phosphate buffer (5:10:3:2:5, v/v/v/v/v), and chromatographed on an Ito Coil Planet Centrifuge in the same system. Active fractions are combined, concentrated and partitioned between methylene chloride and dilute ammonium hydroxide (pH 9.0). The methylene chloride layer is separated and concentrated to yield the desired product as a white foam.

Although the present invention is illustrated in the examples listed above in terms of preferred embodiments, these examples are not to be regarded as limiting the scope of the invention. The above illustrations serve to describe the principles and methodologies involved in creating the types of genetic alterations that can be introduced into *Sac. erythraea* and/or other *Streptomyces* that result in the synthesis of novel glycosylation-modified polyketide products. Although a single Type I alteration, leading to the production of for example, 4"-deoxy-4"-oxo-erythromycin A, is specified herein, it is obvious to those skilled in the art that other Type I changes can be introduced into the *eryB* and/or *eryC* genes leading to novel glycosylation-modified polyketide structures. Examples of additional Type I alterations leading to useful novel compounds include but are not limited to: mutations in the *eryBVII* gene conceivably leading to 3-α-D-mycarosyl-5-ß-D-desosaminoyl-12-hydroxy-erythronolide B and mutations in the *eryCVI* gene conceivably leading to N-3α'-des-dimethyl erythromycin A. Moreover, it is obvious that Type I alterations in two or more different *eryB* and/or *eryC* genes can be combined leading to novel glycosylation-modified polyketide structures. Examples of combinations of two Type I alterations leading to useful compounds include but are not limited to: mutations in the *eryBIV* and *eryBVII* genes conceivably leading to 3-α-D-4"-deoxy-4"-oxo-mycarosyl-5-ß-D-desosaminoyl-12-hydroxy-erythronolide B; mutations in the *eryBIV* and *eryCVI* genes conceivably leading to 4"-deoxy-4"-oxo-(N-3α'-des-dimethyl)-erythromycin A; and mutations in the *eryBIV, eryBVII* , and *eryCVI* genes conceivably leading to 3-α-D-4"-deoxy-4"-oxo-mycarosyl-5-ß-D-(N-3α'-des-dimethyl)-desosaminoyl-12-hydroxy-erythronolide B. All Type I mutations or combinations of two or more Type I mutations in the *eryBII, eryBIV, eryBV, eryBVI, eryBVII, eryCII, eryCIII, eryCIV, eryCV,* or *eryCVI* genes, the *Sac. erythraea* strains that carry said mutations or combinations of mutations, and the corresponding polyketides produced from said strains, therefore, are included within the scope of the present invention.

Although the Type II mutation specified herein was constructed with the *eryBVII* gene on a self-replicating plasmid it is obvious that other *eryB* genes and *eryC* genes can be expressed in an antisense orientation leading to novel glycosylation-modified polyketide structures. Examples of additional Type II alterations leading to useful compounds include but are not limited to: antisense expression of the *eryBIV gene* conceivably leading to 4"-

deoxy-4"-oxo-erythromycin A and antisense expression of the *eryCVI* gene conceivably leading to N-3α'-des-dimethyl erythromycin A. Moreover, it will occur to those skilled in the art that promoters other than the *ermE** promoter, for example the *melC* promoter of pIJ702, will be suitable for antisense expression, and that many self-replicating vectors in addition to pWHM4 will function to carry the antisense alteration. It will also occur to those skilled in the art that a self-replicating vector is not required for this invention and that the antisense alteration can be introduced directly into the chromosome using the same principles employed to construct a Type I gene alteration. An example of a Type II alteration that is introduced directly into the chromosome is the *eryBVII* antisense alteration described in Example 2 wherein DNA segments immediately upstream of the *eryK* gene are used to flank the *ermE-eryBVII*-phage fd terminator grouping in a pWHM3 vector, and this vector is integrated into and then resolved from the chromosome leaving the *ermE*-eryBVII*-phage fd terminator grouping stably incorporated into this nonessential region of the chromosome of *Sac. erythraea* conceivably leading to the production of 3-α-D-mycarosyl-5-ß-D-desosaminoyl-12-hydroxy-erythronolide B. All Type II mutations in the *eryBII, eryBIV, eryBV, eryBVI, eryBVII, eryCII, eryCIII, eryCIV, eryCV,* or *eryCVI* genes whether carried on a self-replicating plasmid or integrated into a nonessential region of the chromosome, the *Sac. erythraea* strains that carry said mutations, and the corresponding polyketides produced from said strains, therefore, are included within the scope of the present invention.

Although Type III alterations, leading to the production of 5-des-chalcosyl-5-desosaminoyl lankamycin in *Streptomyces violaceoniger* and 3-des-oleandrosyl-3-mycarosyl oleandomycin in *Streptomyces antibioticus*, are specified herein, it is obvious that Type III alterations can be introduced into any polyketide producing microorganism leading to novel glycosylation modified polyketides. It will also occur to those skilled in the art that both the *eryB* and *eryC* genes can either be cotransformed into a polyketide producing microorganism or grouped together on a single vector that is introduced into a polyketide producing microorganism. An example of a Type III change using both the *eryB* and *eryC* genes together is their introduction into *Streptomyces violaceoniger* conceivably leading to 3-des-(4"-O-acetylarcanosyl)-3-mycarosyl-5-des-chalcosyl-5-desosaminoyl lankamycin. Although the Type III alterations specified herein have indicated a specific genetic order of the *eryB* or *eryC* genes, it will occur to those skilled at the art that many different genetic arrangements of the *eryB* or *eryC* genes will produce similar results. It will also that occur to those skilled at the art that certain arrangements of the *eryB* and/or *eryC* genes that lack one or more of the respective *eryB* and/or *eryC* genes will lead to the production of novel glycosylated polyketides in which intermediate compounds in the biosynthesis of mycarose and/or desosamine, respectively, such as those outlined in FIGS. 2 and 3, are attached to the polyketide. An example of a Type III alteration in which only a subset of the *eryB* and/or

eryC genes are used is the introduction of a pXC4 derivative that lacks the *eryCVI* gene, removed by digestion of plasmid pXC4 with *Afl*II and *Avr*II followed by treatment with the Klenow fragment of DNA polymerase I and religation, into *Streptomyces violaceoniger* leading to the production of to 5-des-chalcosyl-5-(N-3α'-des-dimethyl desosaminoyl) lankamycin. It will also that occur to those skilled at the art that promoters other than *ermE* or *ermE**, such as the *melC* promoter of plasmid pIJ702, and vectors other than pWHM4 or pIJ702 can also be utilized in the construction of a Type III alteration, and these variants are, of course, considered to be within the scope of the invention. Finally, it will also occur to those skilled in the art that a self-replicating vector is not required for this invention and that an assembly of sugar biosynthesis genes can be introduced directly into the chromosome of a heterologous host using the same principles employed to construct a Type I gene alteration once a nonessential region of the heterologous host chromosome has been identified. Alternatively, plasmids or bacteriophages which undergo site-specific recombination with host genes may also be used to introduce *eryB* and *eryC* genes into a host to effect Type III alterations. All Type III alterations using one or more of the *eryBII, eryBIV, eryBV, eryBVI, eryBVII, eryCII, eryCIII, eryCIV, eryCV,* or *eryCVI* genes, the polyketide producing strains that carry said alterations, and the corresponding polyketides produced from said strains, therefore, are included within the scope of the present invention.

In addition, it is also possible to create combinations of Type I and Type II alterations such that some Type I *eryB* and/or *eryC* mutations are introduced directly into the *Sac. erythraea* chromosome in the appropriate locus, while other *eryB* and/or *eryC* genes are inactivated by Type II alterations using a self-replicating or integrating vector. For example, combination of a Type I alteration, such as a mutation in *eryBIV*, and a Type II alteration, such as transformation with pAS*BVII*, will conceivably lead to production of 3-α-D-4"-deoxy-4"-oxo-mycarosyl-5-B-D-desosaminoyl-12-hydroxy-erythronolide B. All combinations of two or more alterations of Type I and Type II, the *Sac. erythraea* strains that carry such alterations, and the glycosylated polyketides produced from such strains are included within the scope of the present invention.

As an extension of the examples reported with the *eryB* and/or *eryC* genes, it is possible to apply the method described herein to heterologous sugar biosynthesis genes that are similar to the *eryB* and/or *eryC* genes. The construction of strains carrying heterologous sugar biosynthesis genes that lead to the production of novel glycosylated polyketides requires: (i) cloning of the sugar biosynthesis genes from any other glycosylated-polyketide producing actinomycete, (ii) determining the nucleotide sequence of the cloned gene(s); (iii) excising and assembling the cloned gene(s) into vectors suitable for Type I, Type II, or Type III alterations; and (iv) transformation of polyketide producing microorganisms and screening for the novel compound. Any polyketide-associated sugar biosynthesis gene can thus be

precisely excised from the genome of a glycosylated polyketide producing microorganism and altered or arranged with other sugar biosynthesis genes and then introduced into the same or another polyketide producing microorganism to create a novel glycosylated polyketide of predicted structure. Thus, for example, a Type I or Type II alteration of a heterologous gene that is similar to an *eryB* and/or *eryC* gene, such as can be found in the *eryBVII* homolog for the synthesis of L-oleandrose in *Streptomyces antibioticus*, to result in the production of 3-des-L-oleandrosyl-3-D-oleandrosyl oleandomycin is included within the scope of the present invention. Similarly, a Type III assembly of the genes for the synthesis of a sugar other than mycarose or desosamine, such as can be found in the genes for the synthesis of angolosamine in *Streptomyces eurythermus*, and their transformation into *Sac. erythraea* to result in the synthesis of 5-des-desosaminoyl-5-angolosaminoyl-erythromycin A is included within the scope of the present invention.

It will occur to those skilled in the art that the Type I, Type II, and Type III genetic manipulations described herein and the polyketide producing microorganisms into which they are introduced are in no way exclusive. Hence, the choice of a convenient host and the choice of a Type I, Type II, or Type III alteration is based solely on the relatedness of the desired novel glycosylated polyketide to a natural counterpart. Therefore, Type I, Type II, and Type III alterations can be constructed in any polyketide producing microorganism employing either endogenous or exogenous sugar biosynthesis genes. Thus all Type I, Type II, and Type III mutations or various combinations thereof constructed in any polyketide producing microorganism according to the principles described herein, and the respective polyketides produced from such strains, are included within the scope of the present invention. Examples of glycosylated polyketides that can be altered by creating Type I, Type II, or Type III changes in the producing microorganisms include, but are not limited to macrolide antibiotics such as erythromycin, tylosin, spiramycin, etc; aromatic polyketides such as daunorubicin and doxorubicin, etc; polyenes such as candicidin, amphotericins, etc; and other complex polyketides such as avermectin.

Whereas the novel derivatives or modifications of erythromycin described herein have been specified as the A derivatives, such as 4"-deoxy-4"-oxo-erythromycin A, those skilled in the art understand that the wild type strain of *Sac. erythraea* produces a family of erythromycin compounds, including erythromycin A, erythromycin B, erythromycin C, and erythromycin D. Thus, modified strains of *Sac. erythraea*, such as strain ERBIV, for example, would be expected to produce the corresponding members of the 4"-deoxy-4"-oxo-erythromycin family, including 4"-deoxy-4"-oxo-erythromycin A, 4"-deoxy-4"-oxo-erythromycin B, 4"-deoxy-4"-oxo-erythromycin C, and 4"-deoxy-4"-oxo-erythromycin D. Similarly, all other modified strains of *Sac. erythraea* that produce novel glycosylated erythromycin derivatives would be expected to produce the A, B, C, and D forms of said

derivatives. For example, modified *Sac. erythraea* strains that produce 6-deoxyerythromycin, 6,12-dideoxyerythromycin and 6,7-anhydroerythromycin would be expected to produce novel glycosylation-modified polyketides by introduction of the additional modification of a Type I, II or III change in a sugar biosynthesis gene. Therefore, all members of the family of each of the novel erythromycins described herein or produced by these methods are included within the scope of the present invention.

Variations and modifications of the methods for obtaining the desired plasmids, hosts for cloning and choices of vectors and *eryB* and/or *eryC* genes to clone and modify, other than those described herein will occur to those skilled in the art. For example, although we have described the use of plasmids pWHM3, pWHM4, and pIJ702, other vectors can be employed wherein all or part of said plasmids is replaced by other DNA segments that function in a similar manner, such as replacing the pUC19 component of pWHM3 and pWHM4 with pBR322, available from BRL; or employing different segments of the pIJ101 replicon in pWHM3 and pIJ702, or the pJV1 replicon in pWHM4, respectively; or employing selectable markers other than thiostrepton- or ampicillin-resistance. These are just a few of a long list of possible examples all of which are included within the scope of the present invention. Similarly, the segments of the *eryB* and *eryC* loci that have been specified herein to generate the various Type I, Type II, and Type III alterations can readily be substituted for other segments of different length encoding the same functions, either produced by PCR-amplification of genomic DNA or of an isolated clone, or by isolating suitable restriction fragments from *Sac. erythraea*. In the same way it is possible to create Type I mutations functionally equivalent to those described herein by altering through deletion, insertion, or site directed mutagenesis different portions of the corresponding genes. It is also possible to create Type II mutations functionally equivalent to those described herein by employing larger or smaller portions of the corresponding genes; and it is possible to create Type III mutations using larger or smaller segments of the corresponding genes in the same or different linear order described herein. Additional modifications include changes in the restriction sites used for cloning or in the general methodologies described above. All such changes are included in the scope of the present invention. It will also occur to those skilled in the art that different methods are available to ferment *Sac. erythraea* and other polyketide producing microorganisms and to extract the novel polyketides specified herein, and all such methods are also included within the scope of this invention.

It will also be apparent that many modifications and variations of the invention as set forth herein are possible without departing from the spirit and scope thereof, and that, accordingly, such limitations are imposed only as indicated by the appended claims.

We claim:

1. An isolated single or double stranded polynucleotide having a nucleotide sequence which comprises (a) a nucleotide sequence selected from the group consisting of (i) the sense sequence of SEQ ID NO:1 from about nucleotide position 54 to about nucleotide position 1136; (ii) the sense sequence of SEQ ID NO:1 from about nucleotide position 1147 to about nucleotide position 2412; (iii) sense sequence of SEQ ID NO:1 from about nucleotide position 2409 to about nucleotide position 3410; (iv) the sense sequence of SEQ ID NO:2 from about nucleotide position 80 to about nucleotide position 1048; (v) the sense sequence of SEQ ID NO:2 from about nucleotide position 1048 to about nucleotide position 2295; (vi) the sense sequence of SEQ ID NO:2 from about nucleotide position 2348 to about nucleotide position 3061; (vii) the sense sequence of SEQ ID NO:2 from about nucleotide position 3214 to about nucleotide position 4677; (viii) the sense sequence of SEQ ID NO:2 from about nucleotide position 4674 to about nucleotide position 5879; (iv) the sense sequence of SEQ ID NO:2 from about nucleotide position 5917 to about nucleotide position 7386; and (x) the sense sequence of SEQ ID NO:2 from about nucleotide position 7415 to about nucleotide position 7996;

(b) sequences complementary to the sequences of (a);

(c) sequences that, on expression, encode a polypeptide encoded by the sequences of (a); and

(d) analogous sequences that hybridize under stringent conditions to the sequences of (a).

2. The polynucleotide of claim 1 that is a DNA molecule or RNA molecule.

3. The polynucleotide of claim 2 wherein the nucleotide sequence is the nucleotide sequence of (a) selected from the group consisting of (i) the sense sequence of SEQ ID NO:1 from about nucleotide position 54 to about nucleotide position 1136; (ii) the sense sequence of SEQ ID NO:1 from about nucleotide position 1147 to about nucleotide position 2412; (iii) the sense sequence of SEQ ID NO:2 from about nucleotide position 2348 to about nucleotide position 3061; (iv) the sense sequence of SEQ ID NO:2 from about nucleotide position 4674 to about nucleotide position 5879; and (v) the sense sequence of SEQ ID NO:2 from about nucleotide position 5917 to about nucleotide position 7386.

4. The polynucleotide of claim 2 wherein the nucleotide sequence is the nucleotide sequence of (a) selected from the group consisting of (i) sense sequence of SEQ ID NO:1 from about nucleotide position 2409 to about nucleotide position 3410; (ii) the sense

sequence of SEQ ID NO:2 from about nucleotide position 80 to about nucleotide position 1048; (iii) the sense sequence of SEQ ID NO:2 from about nucleotide position 1048 to about nucleotide position 2295; (iv) the sense sequence of SEQ ID NO:2 from about nucleotide position 3214 to about nucleotide position 4677; and (v) the sense sequence of SEQ ID NO:2 from about nucleotide position 7415 to about nucleotide position 7996.

5. The polynucleotide of claim 2 wherein the nucleotide sequence is the nucleotide sequence of (a) having the sense sequence of SEQ ID NO:2 from about nucleotide position 80 to about nucleotide position 1048.

6. A vector comprising the DNA molecule of claim 2.

7. The vector of claim 6 further comprising an enhancer-promoter operatively linked to the polynucleotide.

8. The vector of claim 6 wherein the polynucleotide has the nucleotide sequence of claim 5.

9. A host cell transformed with the vector of claim 6 or claim 7 or claim 8.

10. The transformed host cell of claim 9 that is a bacterial cell.

11. The transformed host cell of claim 10 wherein the bacterial cell is selected from the group consisting of *Streptomyces* and *E. coli*.

12. A method for directing the biosynthesis of specific glycosylation-modified polyketides by genetic manipulation of a polyketide-producing microorganism, said method comprising the steps of:

(1) isolating a sugar biosynthesis gene-containing DNA sequence according to claim 1;

(2) identifying within said gene-containing DNA sequence one or more DNA fragments responsible for the biosynthesis of a polyketide-associated sugar or its attachment to a polyketide;

(3) creating one or more specified changes into said DNA fragment or fragments, thereby resulting in an altered DNA sequence;

(4) introducing said altered DNA sequence into a polyketide-producing microorganism to replace the original sequence, said altered DNA sequence, when translated,

resulting in altered enzymatic activity capable of effecting the production of said specific glycosylation-modified polyketide;

(5) growing a culture of said altered polyketide-producing microorganism under conditions suitable for the formation of said specific glycosylation-modified polyketide; and

(6) isolating said specific glycosylation-modified polyketide from said culture.

13. The method of claim 12 wherein said specified change in said DNA fragment or fragments results in the inactivation of at least one enzymatic activity involved in the biosynthesis of a polyketide-associated sugar or in its attachment to a polyketide.

14. The method of claim 13 wherein said polyketide-associated sugar is L-mycarose.

15. The method of claim 13 wherein said polyketide-associated sugar is D-desosamine.

16. A method for directing the biosynthesis of specific glycosylation-modified polyketides by genetic manipulation of a polyketide-producing microorganism, said method comprising the steps of:

(1) isolating a sugar biosynthesis gene-containing DNA sequence according to claim 1;

(2) identifying within said gene-containing DNA sequence one or more DNA fragments responsible for the biosynthesis of a polyketide-associated sugar or its attachment to a polyketide;

(3) reversing the strand orientation of said DNA fragment or fragments, thereby resulting in an altered DNA sequence which, when transcribed, results in production of an antisense mRNA;

(4) introducing said altered DNA sequence into a polyketide-producing microorganism having an mRNA capable of binding to said antisense mRNA to produce an altered polyketide-producing microorganism capable of producing said specific glycosylation-modified polyketide;

(5) growing a culture of said altered polyketide-producing microorganism under conditions suitable for the formation of said specific glycosylation-modified polyketide; and

(6) isolating said specific glycosylation-modified polyketide from said culture.

17. A method for directing the biosynthesis of specific glycosylation-modified polyketides by genetic manipulation of a polyketide-producing microorganism, said method comprising the steps of:

(1) isolating a sugar biosynthesis gene-containing DNA sequence according to claim

1;

(2) identifying within said gene-containing DNA sequence one or more DNA fragments responsible for the biosynthesis of a polyketide-associated sugar or its attachment to a polyketide;

(3) introducing said DNA fragment or fragments into a distinct polyketide-producing microorganism to produce an altered polyketide-producing microorganism capable of producing said specific glycosylation-modified polyketide;

(4) growing a culture of said polyketide-producing microorganism containing said DNA fragment or fragments under conditions suitable for the formation of said specific glycosylation-modified polyketide; and

(6) isolating said specific glycosylation-modified polyketide from said culture.

18. The method of claim 13 or claim 16 or claim 17 wherein said DNA fragment comprises one or more genes which encode an enzymatic activity involved in the biosynthesis of L-mycarose or in its attachment to a polyketide.

19. The method of claim 13 or claim 16 or claim 17 wherein said DNA fragment comprises one or more genes which encode an enzymatic activity involved in the biosynthesis of D-desosamine or in its attachment to a polyketide.

20. The method of claim 13 or claim 16 or claim 17 wherein said DNA fragment is the sequence of claim 8.

21. An isolated polypeptide having an amino acid sequence encoded by a nucleotide sequence selected from the group consisting of the sense sequence of SEQ ID NO:1 from about nucleotide position 54 to about nucleotide position 1136; the sense sequence of SEQ ID NO:1 from about nucleotide position 1147 to about nucleotide position 2412; sense sequence of SEQ ID NO:1 from about nucleotide position 2409 to about nucleotide position 3410; the sense sequence of SEQ ID NO:2 from about nucleotide position 80 to about nucleotide position 1048; the sense sequence of SEQ ID NO:2 from about nucleotide position 1048 to about nucleotide position 2295; the sense sequence of SEQ ID NO:2 from about nucleotide position 2348 to about nucleotide position 3061; the sense sequence of SEQ ID NO:2 from about nucleotide position 3214 to about nucleotide position 4677 ; the sense sequence of SEQ ID NO:2 from about nucleotide position 4674 to about nucleotide position 5879; the sense sequence of SEQ ID NO:2 from about nucleotide position 5917 to about nucleotide position 7386; and the sense sequence of SEQ ID NO:2 from about nucleotide position 7415 to about nucleotide position 7996.

22. An isolated polypeptide of claim 31 encoded by the sequence of SEQ ID NO:2 from about nucleotide position 80 to about nucleotide position 1048.


K
BVII
CV
CIV
BVI
CVI
BV
BIV
AI
AII
AIII
CII
CIII
BII
G
F
BIII
BI
ermE
CI

FIG. 1A

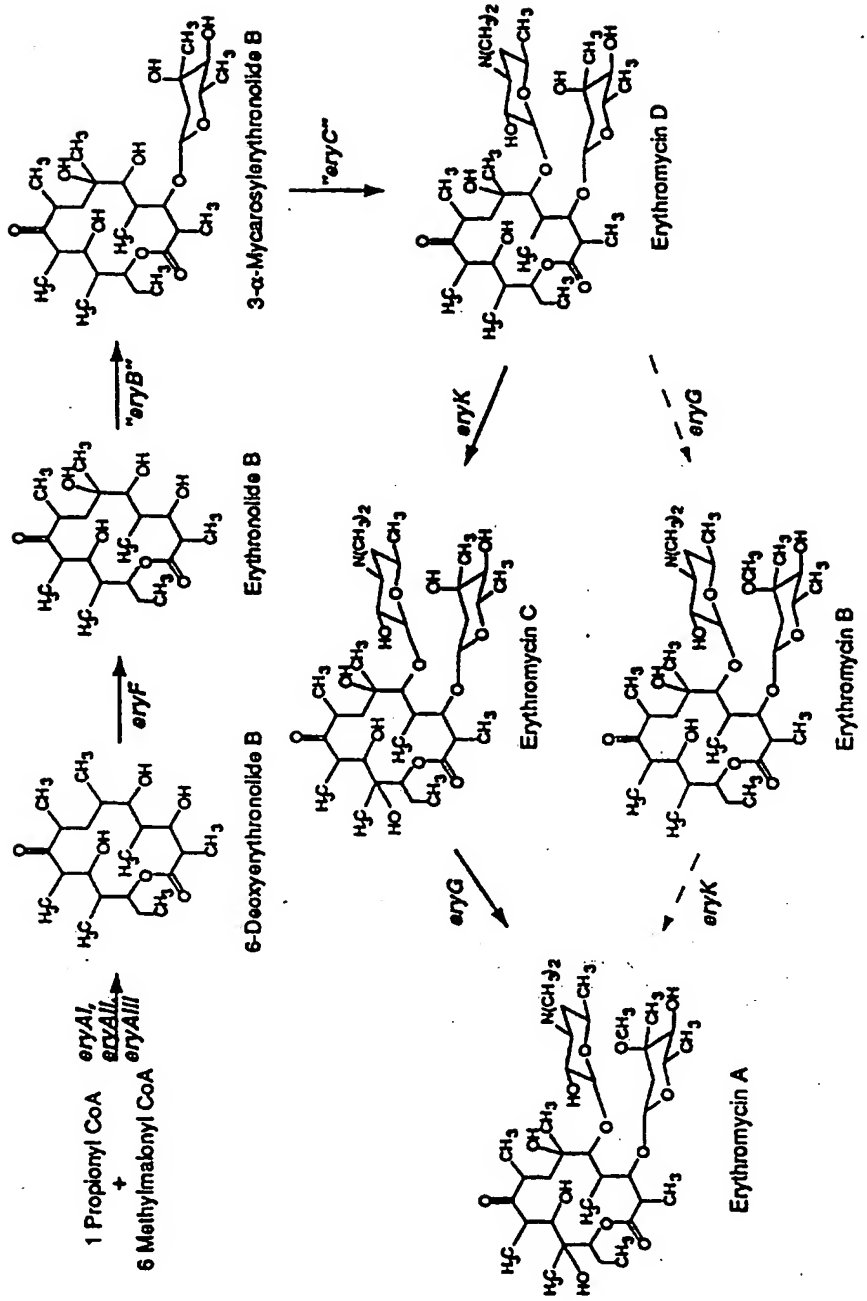
1 Propionyl CoA
+
6 Methylmalonyl CoA
eryAI,
eryAII,
eryAIII
eryF
6-Deoxyerythronolide B
"eryB"
Erythronolide B
"eryC"
3-α-Mycarosylerythronolide B
eryK
eryG
Erythromycin C
Erythromycin D
eryG
eryK
Erythromycin A
Erythromycin B

FIG. 1B

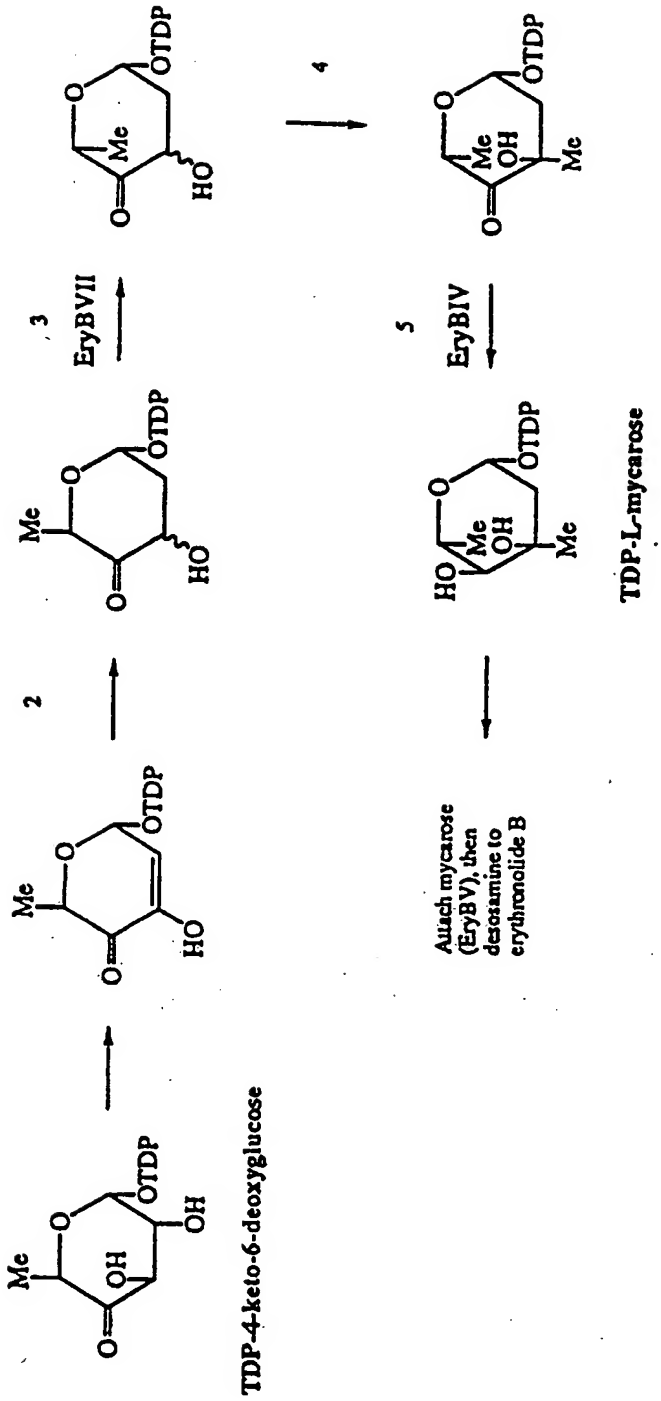
TDP-4-keto-6-deoxyglucose
2
3
EryBVII
4
5
EryBIV
Attach mycarose (EryBV), then desosamine to erythronolide B
TDP-L-mycarose

FIG. 2

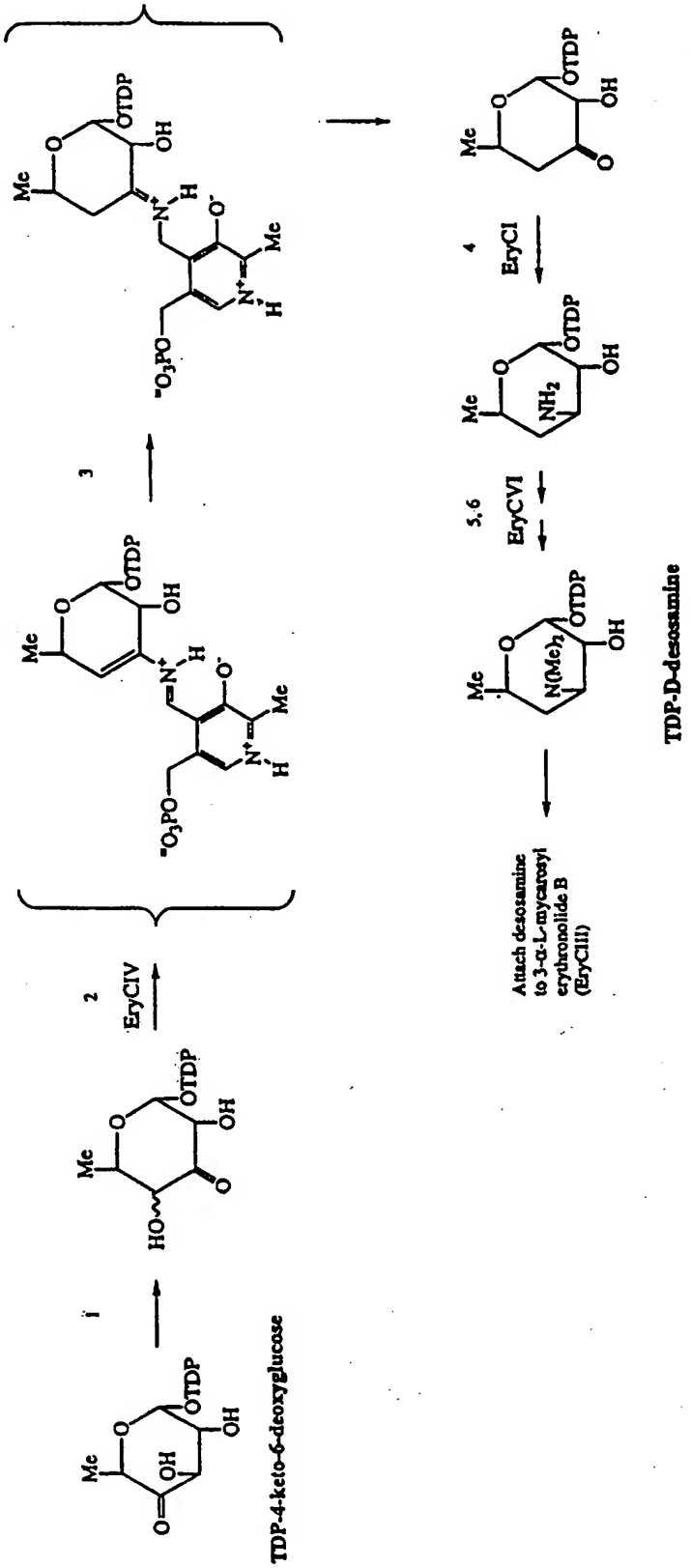
TDP-4-keto-6-deoxyglucose
1
2
EryCIV
3
4
EryCI
5,6
EryCVI
Attach desosamine to 3-α-L-mycarosyl erythronolide B (EryCIII)
TDP-D-desosamine

FIG. 3

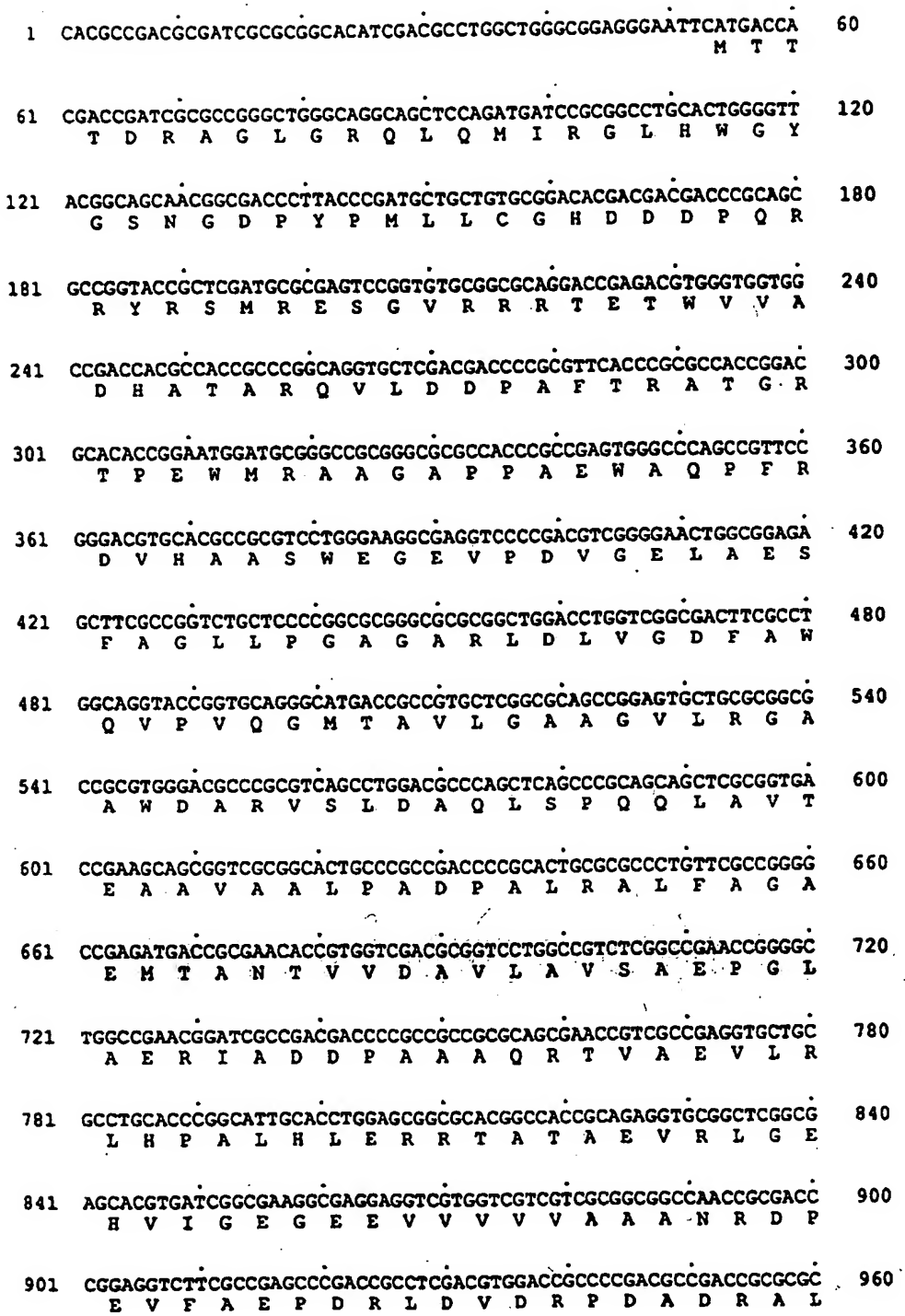
1 CACGCCGACGCGATCGCGCGGCACATCGACGCCTGGCTGGGCGGAGGGAATTCATGACCA 60
M T T
61 CGACCGATCGCGCCGGGCTGGGCAGGCAGCTCCAGATGATCCGCGGCCTGCACTGGGGTT 120
T D R A G L G R Q L Q M I R G L H W G Y
121 ACGGCAGCAACGGCGACCCTTACCCGATGCTGCTGTGCGGACACGACGACGACCCGCAGC 180
G S N G D P Y P M L L C G H D D D P Q R
181 GCCGGTACCGCTCGATGCGCGAGTCCGGTGTGCGGCGCAGGACCGAGACGTGGGTGGTGG 240
R Y R S M R E S G V R R R T E T W V V A
241 CCGACCACGCCACCGCCCGGCAGGTGCTCGACGACCCCGCGTTCACCCGCGCCACCGGAC 300
D H A T A R Q V L D D P A F T R A T G R
301 GCACACCGGAATGGATGCGGGCCGCGGGCGCGCCACCCGCCGAGTGGGCCCAGCCGTTCC 360
T P E W M R A A G A P P A E W A Q P F R
361 GGGACGTGCACGCCGCGTCCTGGGAAGGCGAGGTCCCCGACGTCGGGGAACTGGCGGAGA 420
D V H A A S W E G E V P D V G E L A E S
421 GCTTCGCCGGTCTGCTCCCCGGCGCGGGCGCGCGGCTGGACCTGGTCGGCGACTTCGCCT 480
F A G L L P G A G A R L D L V G D F A W
481 GGCAGGTACCGGTGCAGGGCATGACCGCCGTGCTCGGCGCAGCCGGAGTGCTGCGCGGCG 540
Q V P V Q G M T A V L G A A G V L R G A
541 CCGCGTGGGACGCCCGCGTCAGCCTGGACGCCCAGCTCAGCCCGCAGCAGCTCGCGGTGA 600
A W D A R V S L D A Q L S P Q Q L A V T
601 CCGAAGCAGCGGTCGCGGCACTGCCCGCCGACCCCGCACTGCGCGCCCTGTTCGCCGGGG 660
E A A V A A L P A D P A L R A L F A G A
661 CCGAGATGACCGCGAACACCGTGGTCGACGCGGTCCTGGCCGTCTCGGCCGAACCGGGGC 720
E M T A N T V V D A V L A V S A E P G L
721 TGGCCGAACGGATCGCCGACGACCCCGCCGCCGCGCAGCGAACCGTCGCCGAGGTGCTGC 780
A E R I A D D P A A A Q R T V A E V L R
781 GCCTGCACCCGGCATTGCACCTGGAGCGGCGCACGGCCACCGCAGAGGTGCGGCTCGGCG 840
L H P A L H L E R R T A T A E V R L G E
841 AGCACGTGATCGGCGAAGGCGAGGAGGTCGTGGTCGTCGTCGCGGCGGCCAACCGCGACC 900
H V I G E G E E V V V V V A A A N R D P
901 CGGAGGTCTTCGCCGAGCCCGACCGCCTCGACGTGGACCGCCCCGACGCCGACCGCGCGC 960
E V F A E P D R L D V D R P D A D R A L

FIG. 4A-1

```
961  TGTCGGCACATCGCGGCCACCCCGGCAGGCTGGAGGAGCTGGTCACCGCGCTCGCCACCG  1020
       S  A  H  R  G  H  P  G  R  L  E  E  L  V  T  A  L  A  T  A

1021 CCGCACTGCGGGCCGCGGCCAAGGCGCTGCCCGGACTCACGCCCAGCGGCCCGGTCGTCC  1080
       A  L  R  A  A  A  K  A  L  P  G  L  T  P  S  G  P  V  V  R

1081 GGCGCCGCCGATCACCCGTCCTGCGGGGAACCAACCGCTGCCCCGTCGAGCTCTGAGGAT  1140
       R  R  R  S  P  V  L  R  G  T  N  R  C  P  V  E  L  *

1141 TCCGCGATGCGCGTCGTCTTCTCCTCCATGGCCAGCAAGAGCCACCTCTTCGGCCTCGTC  1200
          M  R  V  V  F  S  S  M  A  S  K  S  H  L  F  G  L  V

1201 CCCCTCGCATGGGCGTTCCGCGCGGCGGGGCACGAGGTCCGCGTGGTCGCGTCCCCGGCG  1260
       P  L  A  W  A  F  R  A  A  G  H  E  V  R  V  V  A  S  P  A

1261 CTCACCGAGGACATCACCGCGGCCGGGCTGACCGCCGTCCCGGTCGGCACCGACGTCGAC  1320
       L  T  E  D  I  T  A  A  G  L  T  A  V  P  V  G  T  D  V  D

1321 CTCGTGGACTTCATGACCCACGCGGGCCACGACATCATCGACTACGTCCGGAGCCTGGAC  1380
       L  V  D  F  M  T  H  A  G  H  D  I  I  D  Y  V  R  S  L  D

1381 TTCAGCGAGCGGGACCCCGCCACCTTGACCTGGGAGCACCTGCGGGGCATGCAGACCGTG  1440
       F  S  E  R  D  P  A  T  L  T  W  E  H  L  R  G  M  Q  T  V

1441 CTCACCCCGACCTTCTACGCCCTGATGAGCCCGGACACGCTCATCGAAGGCATGGTCTCG  1500
       L  T  P  T  F  Y  A  L  M  S  P  D  T  L  I  E  G  M  V  S

1501 TTCTGCCGGAAGTGGCGGCCCGACCTGGTCATCTGGGAGCCGCTCACCTTCGCCGCGCCC  1560
       F  C  R  K  W  R  P  D  L  V  I  W  E  P  L  T  F  A  A  P

1561 ATCGCGGGCGCGGTGACCGGAACGCCGCACGCGCGGCTGCTGTGGGGACCCGACATCACC  1620
       I  A  G  A  V  T  G  T  P  H  A  R  L  L  W  G  P  D  I  T

1621 ACCCGGGCGCGGCAGAACTTCCTCGGCCTGCTGCCCGACCAGCCGGAGGAGCACCGGGAG  1680
       T  R  A  R  Q  N  F  L  G  L  L  P  D  Q  P  E  E  H  R  E

1681 GGCCCGCTCGCCGAGTGGCTCACCTGGACGCTGGAGAAGTACGGCGGCCCGGCCTTCGAC  1740
       G  P  L  A  E  W  L  T  W  T  L  E  K  Y  G  G  P  A  F  D

1741 GAGGAGGTGGTCGTCGGGCAGTGGACGATCGACCCCGCCCCGGCCGCGATCAGGCTCGAC  1800
       E  E  V  V  V  G  Q  W  T  I  D  P  A  P  A  A  I  R  L  D

1801 ACCGGCCTGAAGACCGTCGGGATGCGCTACGTCGACTACAACGGGCCGTCCGTGGTGCCG  1860
       T  G  L  K  T  V  G  M  R  Y  V  D  Y  N  G  P  S  V  V  P

1861 GAATGGCTGCACGACGAGCCCGAGCGCCGCCGCGTGTGCCTCACGCTCGGGATCTCCAGC  1920
       E  W  L  H  D  E  P  E  R  R  R  V  C  L  T  L  G  I  S  S
```

FIG. 4A–2

```
1921  CGCGAGAACAGCATCGGGCAGGTCTCCATCGAGGAGCTGCTGGGTGCCGTCGGCGACGTC  1980
      R  E  N  S  I  G  Q  V  S  I  E  E  L  L  G  A  V  G  D  V

1981  GACGCCGAGATCATCGCGACCTTCGACGCGCAGCAGCTAGAAGGCGTCGCGAACATCCCG  2040
      D  A  E  I  I  A  T  F  D  A  Q  Q  L  E  G  V  A  N  I  P

2041  CACAACGTCCGCACGGTCGGCTTCGTCCCGATGCACGCGCTGCTGCCGACCTGCGCGGCG  2100
      H  N  V  R  T  V  G  F  V  P  M  H  A  L  L  P  T  C  A  A

2101  ACGGTGCACCACGGCGGACCCGGGAGCTGGCACACCGCGGCGATCCACGGCGTGCCGCAG  2160
      T  V  H  H  G  G  P  G  S  W  H  T  A  A  I  H  G  V  P  Q

2161  GTGATCCTGCCCGACGGCTGGGACACCGGCGTGCGCGCGCAGCGCACGCAGGAATTCGGG  2220
      V  I  L  P  D  G  W  D  T  G  V  R  A  Q  R  T  Q  E  F  G

2221  GCGGGGATCGCGCTGCCCGTGCCCGAGCTGACCCCCGACCAGCTCCGGGAGTCGGTGAAG  2280
      A  G  I  A  L  P  V  P  E  L  T  P  D  Q  L  R  E  S  V  K

2281  CGGGTCCTCGACGACCCGGCCCACCGCGCCGGCGCGGCGCGGATGCGCGACGACATGCTC  2340
      R  V  L  D  D  P  A  H  R  A  G  A  A  R  M  R  D  D  M  L

2341  GCGGAGCCGTCACCGGCCGAGGTCGTCGGCATCTGCGAGGAACTGGCCGCAGGAAGGAGA  2400
      A  E  P  S  P  A  E  V  V  G  I  C  E  E  L  A  A  G  R  R

2401  GAACCACGATGACCACCGACGCCGCGACGCACGTGCGGCTCGGGCGTTCCGCGCTGCTCA  2460
      E  P  R  *
               M  T  T  D  A  A  T  H  V  R  L  G  R  S  A  L  L  T

2461  CCAGCAGGCTCTGGCTCGGCACGGTGAACTTCAGCGGACGCGTCGAGGACGACGACGCGC  2520
        S  R  L  W  L  G  T  V  N  F  S  G  R  V  E  D  D  D  A  L

2521  TGCGCCTGATGGACCACGCCCGGGACCGCGGCATCAACTGCCTCGACACCGCCGACATGT  2580
        R  L  M  D  H  A  R  D  R  G  I  N  C  L  D  T  A  D  M  Y

2581  ACGGCTGGCGGCTCTACAAGGGCCACACCGAGGAGCTGGTGGGCAGGTGGCTGGCCCAGG  2640
        G  W  R  L  Y  K  G  H  T  E  E  L  V  G  R  W  L  A  Q  G

2641  GCGGCGGACGGCGCGAGGACACCGTGCTGGCGACCAAGGTCGGCGGCGAGATGAGCGAGC  2700
        G  G  R  R  E  D  T  V  L  A  T  K  V  G  G  E  M  S  E  R

2701  GCGTCAACGACAGCGGGCTGTCGGCGCGGCACATCATCGCCTCCTGCGAGGGATCGCTGC  2760
        V  N  D  S  G  L  S  A  R  H  I  I  A  S  C  E  G  S  L  R

2761  GCAGGCTGGGCGTCGACCACATCGACGTCTACCAGATGCACCACATCGACCGGTCCGCGC  2820
        R  L  G  V  D  H  I  D  V  Y  Q  M  H  H  I  D  R  S  A  P

2821  CGTGGGACGAGGTGTGGCAGGCCATGGACAGCCTCGTCGCCAGCGGCAAGGTCTCCTACG  2880
        W  D  E  V  W  Q  A  M  D  S  L  V  A  S  G  K  V  S  Y  V
```

FIG. 4A–3

```
2881  TCGGCTCGTCGAACTTCGCGGGCTGGCACATCGCCGCCGCGCAGGAGAACGCCGCCCGCC  2940
        G  S  S  N  F  A  G  W  H  I  A  A  A  Q  E  N  A  A  R  R

2941  GCCACTCCCTGGGCATGGTCTCCCACCAGTGCCTGTACAACCTGGCGGTCCGGCACGCCG  3000
        H  S  L  G  M  V  S  H  Q  C  L  Y  N  L  A  V  R  H  A  E

3001  AGCTGGAGGTGCTGCCCGCCGCGCAGGCCTACGGGCTCGGCGTCTTCGCCTGGTCGCCGC  3060
        L  E  V  L  P  A  A  Q  A  Y  G  L  G  V  F  A  W  S  P  L

3061  TGCACGGCGGCCTGCTCAGCGGAGCGCTGGAGAAGCTGGCCGCGGGCACCGCGGTGAAGT  3120
        H  G  G  L  L  S  G  A  L  E  K  L  A  A  G  T  A  V  K  S

3121  CGGCGCAGGGCCGTGCGCAGGTGCTGTTGCCGTCCCTGCGCCCGGCGATCGAGGCCTACG  3180
        A  Q  G  R  A  Q  V  L  L  P  S  L  R  P  A  I  E  A  Y  E

3181  AGAAGTTCTGCCGCAACCTCGGCGAAGACCCGGCCGAGGTGGGGCTCGCATGGGTGCTGT  3240
        K  F  C  R  N  L  G  E  D  P  A  E  V  G  L  A  W  V  L  S

3241  CCCGGCCCGGCATCGCCGGCGCCGTCATCGGCCCGCGAACCCCCGAGCAGCTCGACTCCG  3300
        R  P  G  I  A  G  A  V  I  G  P  R  T  P  E  Q  L  D  S  A

3301  CGCTGAAGGCGTCCGCGATGACCCTGGACGAGCAGGCGCTGTCCGAACTGGACGAGATCT  3360
        L  K  A  S  A  M  T  L  D  E  Q  A  L  S  E  L  D  E  I  F

3361  TCCCCGCGGTGGCCTCCGGCGGCGCGGCGCCGGAAGCCTGGTTGCAGTGAGCACAAGAGG  3420
        P  A  V  A  S  G  G  A  A  P  E  A  W  L  Q  *

3421  AACCGAGAAAGGATACGGCTGGTGAGCGTGAAGCAGAAGTCAGCGTTGCAGGACCTGGTC  3480

3481  GACTTCGCCAAGTGGCACGTGTGGACCAGGGTGCGGCCGTCCAGCCGTGCGCGCCTGGCC  3540

3541  TACGAGCTGTTCGCCGACGACCACGAGGCCACGACCGAGGGCGCCTACATCAACCTCGGC  3600

3601  TACTGGAAGCCCGGGTGCGCCGGCCTGGAGGAGGCCAACCAGGAGCTGGCGAACCAGCTC  3660

3661  GCCGAGGCCGCGGGGATCAGCGAGGGCGACGAGGTGCTCGACGTCGGGTTCGGGCTCGGC  3720

3721  GCGCAGGACTTCTTCTGGCTCGACCTGCAGCCAGCT  3756
```

FIG. 4A–4

```
1   CGGGTTGCCGCACATCGCGCTGGGGAGATTCTTTGAATTTCGCCCGTAGCACCGACCTGG   60

61  AAAGCGAGCAAATGCTCCGGTGAATGGGATCAGTGATTCCCCGCGTCAATTGATCACCCT  120
                    V  N  G  I  S  D  S  P  R  Q  L  I  T  L

121 TCTGGGCGCTTCCGGCTTCGTCGGGAGCGCGGTTCTGCGCGAGCTGCGCGACCACCCGGT  180
      L  G  A  S  G  F  V  G  S  A  V  L  R  E  L  R  D  H  P  V

181 CCGGCTGCGCGCGGTGTCCCGCGGCGGAGCGCCCGCGGTTCCGCCCGGCGCCGCGGAGGT  240
      R  L  R  A  V  S  R  G  G  A  P  A  V  P  P  G  A  A  E  V

241 CGAGGACCTGCGCGCCGACCTGCTGGAACCGGGCCGGGCCGCCGCCGCGATCGAGGACGC  300
      E  D  L  R  A  D  L  L  E  P  G  R  A  A  A  A  I  E  D  A

301 CGACGTGATCGTGCACCTGGTGGCGCACGCAGCGGGCGGTTCCACCTGGCGCAGCGCCAC  360
      D  V  I  V  H  L  V  A  H  A  A  G  G  S  T  W  R  S  A  T

361 CTCCGACCCGGAAGCCGAGCGGGTCAACGTCGGCCTGATGCACGACCTCGTCGGCGCGCT  420
      S  D  P  E  A  E  R  V  N  V  G  L  M  H  D  L  V  G  A  L

421 GCACGATCGCCGCAGGTCGACGCCGCCCGTGTTGCTCTACGCGAGCACCGCACAGGCCGC  480
      H  D  R  R  R  S  T  P  P  V  L  L  Y  A  S  T  A  Q  A  A

481 GAACCCGTCGGCGGCCAGCAGGTACGCGCAGCAGAAGACCGAGGCCGAGCGCATCCTGCG  540
      N  P  S  A  A  S  R  Y  A  Q  Q  K  T  E  A  E  R  I  L  R

541 CAAAGCCACCGACGAGGGCCGGGTGCGCGGCGTGATCCTGCGGCTGCCCGCGGTCTACGG  600
      K  A  T  D  E  G  R  V  R  G  V  I  L  R  L  P  A  V  Y  G

601 CCAGAGCGGCCCGTCCGGCCCCATGGGGCGGGGCGTGGTCGCAGCGATGATCCGGCGTGC  660
      Q  S  G  P  S  G  P  M  G  R  G  V  V  A  A  M  I  R  R  A

661 CCTCGCCGGCGAGCCGCTCACCATGTGGCACGACGGCGGCGTGCGCCGCGACCTGCTGCA  720
      L  A  G  E  P  L  T  M  W  H  D  G  G  V  R  R  D  L  L  H

721 CGTCGAGGACGTGGCCACCGCGTTCGCCGCCGCGCTGGAGCACCACGACGCGCTGGCCGG  780
      V  E  D  V  A  T  A  F  A  A  A  L  E  H  H  D  A  L  A  G

781 CGGCACGTGGGCGCTGGGCGCCGACCGATCCGAGCCGCTCGGCGACATCTTCCGGGCCGT  840
      G  T  W  A  L  G  A  D  R  S  E  P  L  G  D  I  F  R  A  V

841 CTCCGGCAGCGTCGCCCGGCAGACCGGCAGCCCCGCCGTCGACGTGGTCACCGTGCCCGC  900
      S  G  S  V  A  R  Q  T  G  S  P  A  V  D  V  V  T  V  P  A

901 GCCCGAGCACGCCGAGGCCAACGACTTCCGCAGCGACGACATCGACTCCACCGAGTTCCG  960
      P  E  H  A  E  A  N  D  F  R  S  D  D  I  D  S  T  E  F  R
```

FIG. 4B–1

```
961  CAGCCGGACCGGCTGGCGCCCCCGGGTTTCCCTCACCGACGGCATCGACCGGACGGTGGC  1020
      S  R  T  G  W  R  P  R  V  S  L  T  D  G  I  D  R  T  V  A

1021 CGCCCTGACCCCCACCGAGGAGCACTAGTGCGGGTACTGCTGACGTCCTTCGCGCACCGC  1080
      A  L  T  P  T  E  E  H  *
                                  V  R  V  L  L  T  S  F  A  H  R

1081 ACGCACTTCCAGGGACTGGTCCCGCTGGCGTGGGCGCTGCGCACCGCGGGTCACGACGTG  1140
      T  H  F  Q  G  L  V  P  L  A  W  A  L  R  T  A  G  H  D  V

1141 CGCGTGGCCGCCCAGCCCGCGCTCACCGACGCGGTCATCGGCGCCGGTCTCACCGCGGTA  1200
      R  V  A  A  Q  P  A  L  T  D  A  V  I  G  A  G  L  T  A  V

1201 CCCGTCGGCTCCGACCACCGGCTGTTCGACATCGTCCCGGAAGTCGCCGCTCAGGTGCAC  1260
      P  V  G  S  D  H  R  L  F  D  I  V  P  E  V  A  A  Q  V  H

1261 CGCTACTCCTTCTACCTGGACTTCTACCACCGCGAGCAGGAGCTGCACTCGTGGGAGTTC  1320
      R  Y  S  F  Y  L  D  F  Y  H  R  E  Q  E  L  H  S  W  E  F

1321 CTGCTCGGCATGCAGGAGGCCACCTCGCGGTGGGTATACCCGGTGGTCAACAACGACTCC  1380
      L  L  G  M  Q  E  A  T  S  R  W  V  Y  P  V  V  N  N  D  S

1381 TTCGTCGCCGAGCTGGTCGACTTCGCCCGGGACTGGCGTCCTGACCTGGTGCTCTGGGAG  1440
      F  V  A  E  L  V  D  F  A  R  D  W  R  P  D  L  V  L  W  E

1441 CCGTTCACCTTCGCCGGCGCCGTCGCGGCCCGGGCCTGCGGAGCCGCGCACGCCCGGCTG  1500
      P  F  T  F  A  G  A  V  A  A  R  A  C  G  A  A  H  A  R  L

1501 CTGTGGGGCAGCGACCTCACCGGCTACTTCCGCGGCCGGTTCCAGGCGCAACGCCTGCGA  1560
      L  W  G  S  D  L  T  G  Y  F  R  G  R  F  Q  A  Q  R  L  R

1561 CGGCCGCCGGAGGACCGGCCGGACCCGCTGGGCACGTGGCTGACCGAGGTCGCGGGGCGC  1620
      R  P  P  E  D  R  P  D  P  L  G  T  W  L  T  E  V  A  G  R

1621 TTCGGCGTCGAATTCGGCGAGGACCTCGCGGTCGGGCAGTGGTCGGTCGACCAGTTGCCG  1680
      F  G  V  E  F  G  E  D  L  A  V  G  Q  W  S  V  D  Q  L  P

1681 CCGAGTTTCCGGCTGGACACCGGAATGGAAACCGTTGTCGCGCGGACCCTGCCCTACAAC  1740
      P  S  F  R  L  D  T  G  M  E  T  V  V  A  R  T  L  P  Y  N

1741 GGCGCGTCGGTGGTTCCGGACTGGCTCAAGAAGGGCAGTGCGACTCGACGCATCTGCATT  1800
      G  A  S  V  V  P  D  W  L  K  K  G  S  A  T  R  R  I  C  I

1801 ACCGGAGGGTTCTCCGGACTCGGGCTCGCCGCCGATGCCGATCAGTTCGCGCGGACGCTC  1860
      T  G  G  F  S  G  L  G  L  A  A  D  A  D  Q  F  A  R  T  L
```

FIG. 4B–2

```
1861  GCGCAGCTCGCGCGATTCGATGGCGAAATCGTGGTTACGGGTTCCGGTCCGGATACCTCC  1920
      A  Q  L  A  R  F  D  G  E  I  V  V  T  G  S  G  P  D  T  S

1921  GCGGTACCGGACAACATTCGTTTGGTGGATTTCGTTCCGATGGGCGTTCTGCTCCAGAAC  1980
      A  V  P  D  N  I  R  L  V  D  F  V  P  M  G  V  L  L  Q  N

1981  TGCGCGGCGATCATCCACCACGGCGGGGCCGGAACCTGGGCCACGGCACTGCACCACGGA  2040
      C  A  A  I  I  H  H  G  G  A  G  T  W  A  T  A  L  H  H  G

2041  ATTCCGCAAATATCAGTTGCACATGAATGGGATTGCATGCTACGCGGCCAGCAGACCGCG  2100
      I  P  Q  I  S  V  A  H  E  W  D  C  M  L  R  G  Q  Q  T  A

2101  GAACTGGGCGCGGGAATCTACCTCCGGCCGGACGAGGTCGATGCCGACTCATTGGCGAGC  2160
      E  L  G  A  G  I  Y  L  R  P  D  E  V  D  A  D  S  L  A  S

2161  GCCCTCACCCAGGTGGTCGAGGACCCCACCTACACCGAGAACGCGGTGAAGCTTCGCGAG  2220
      A  L  T  Q  V  V  E  D  P  T  Y  T  E  N  A  V  K  L  R  E

2221  GAGGCGCTGTCCGACCCGACGCCGCAGGAGATCGTCCCGCGACTGGAGGAACTCACGCGC  2280
      E  A  L  S  D  P  T  P  Q  E  I  V  P  R  L  E  E  L  T  R

2281  CGCCACGCCGGCTAGCGGTTTCCGACCGACAAGTCCGTCCGACAGCACACCTCCGGAGGG  2340
      R  H  A  G  *

2341  AGCAGGGATGTACGAGGGCGGGTTCGCCGAGCTTTACGACCGGTTCTACCGCGGCCGGGG  2400
            M  Y  E  G  G  F  A  E  L  Y  D  R  F  Y  R  G  R  G

2401  CAAGGACTACGCGGCCGAGGCCGCGCAGGTCGCGCGGCTGGTCAGAGACCGCCTGCCCTC  2460
      K  D  Y  A  A  E  A  A  Q  V  A  R  L  V  R  D  R  L  P  S

2461  GGCTTCCTCGCTGCTCGACGTGGCCTGCGGGACCGGCACCCACCTGCGCCGGTTCGCCGA  2520
      A  S  S  L  L  D  V  A  C  G  T  G  T  H  L  R  R  F  A  D

2521  CCTCTTCGACGACGTGACCGGGCTGGAGCTGTCGGCGGCGATGATCGAGGTCGCCCGGCC  2580
      L  F  D  D  V  T  G  L  E  L  S  A  A  M  I  E  V  A  R  P

2581  GCAGCTCGGCGGCATCCCGGTGCTGCAGGGCGACATGCGCGACTTCGCGCTGGATCGCGA  2640
      Q  L  G  G  I  P  V  L  Q  G  D  M  R  D  F  A  L  D  R  E

2641  GTTCGACGCCGTCACCTGCATGTTCAGCTCCATCGGGCACATGCGCGACGGCGCCGAGCT  2700
      F  D  A  V  T  C  M  F  S  S  I  G  H  M  R  D  G  A  E  L

2701  GGACCAGGCGCTGGCGTCCTTCGCCCGCCACCTCGCCCCCGGCGGCGTCGTGGTGGTCGA  2760
      D  Q  A  L  A  S  F  A  R  H  L  A  P  G  G  V  V  V  V  E

2761  ACCGTGGTGGTTCCCGGAGGACTTCCTCGACGGCTACGTGGCCGGTGACGTGGTGCGCGA  2820
      P  W  W  F  P  E  D  F  L  D  G  Y  V  A  G  D  V  V  R  D
```

FIG. 4B–3

```
2821  CGGCGACCTGACGATCTCGCGCGTCTCGCACTCCGTGCGCGCCGGCGGCGCGACCCGGAT  2880
        G  D  L  T  I  S  R  V  S  H  S  V  R  A  G  G  A  T  R  M

2881  GGAGATCCACTGGGTCGTGGCCGACGCGGTGAACGGTCCGCGGCACCACGTGGAGCACTA  2940
        E  I  H  W  V  V  A  D  A  V  N  G  P  R  H  H  V  E  H  Y

2941  CGAGATCACGCTCTTCGAGCGGCAGCAGTACGAGAAGGCCTTCACCGCGGCCGGTTGCGC  3000
        E  I  T  L  F  E  R  Q  Q  Y  E  K  A  F  T  A  A  G  C  A

3001  TGTGCAGTACCTGGAGGGCGGACCCTCCGGACGCGGGTTGTTCGTCGGTGTGCGCGGATG  3060
        V  Q  Y  L  E  G  G  P  S  G  R  G  L  F  V  G  V  R  G  *

3061  ACCCGTGCGTCGCGTTTTCCGTTCCTGGCACAGGTGATCCGCTCCACGGGCCCTTTCCCC  3120

3121  GCCGTGACCGGACCCTTACAGTGAGTGCGGGTCTTGATCGACAACGCCCGGCGGCAGCAA  3180

3181  GCGGAGCCGTCGACGACACCGCAGGGAGAGTCGATGGGTGATCGGACCGGCGACCGGACG  3240
                                    M  G  D  R  T  G  D  R  T

3241  ATTCCGGAATCCTCGCAGACCGCAACGCGTTTCCTGCTCGGCGACGGCGGAATCCCCACC  3300
        I  P  E  S  S  Q  T  A  T  R  F  L  L  G  D  G  G  I  P  T

3301  GCCACGGCGGAAACCCACGACTGGCTGACCCGCAACGGCGCCGAGCAGCGGCTCGAGGTG  3360
        A  T  A  E  T  H  D  W  L  T  R  N  G  A  E  Q  R  L  E  V

3361  GCGCGCGTGCCGTTCAGCGCCATGGACCGCTGGTCGTTCCAGCCCGAGGACGGCAGGCTC  3420
        A  R  V  P  F  S  A  M  D  R  W  S  F  Q  P  E  D  G  R  L

3421  GCCCACGAGTCCGGGCGCTTCTTCTCCATCGAGGGCCTGCACGTGCGGACGAACTTCGGC  3480
        A  H  E  S  G  R  F  F  S  I  E  G  L  H  V  R  T  N  F  G

3481  TGGCGGCGGGACTGGATCCAGCCCATCATCGTGCAGCCCGAGATCGGCTTCCTCGGCCTC  3540
        W  R  R  D  W  I  Q  P  I  I  V  Q  P  E  I  G  F  L  G  L

3541  ATCGTCAAGGAGTTCGACGGTGTGCTGCACGTGCTGGCGCAGGCCAAGGCCGAGCCGGGC  3600
        I  V  K  E  F  D  G  V  L  H  V  L  A  Q  A  K  A  E  P  G

3601  AACATCAACGCCGTCCAGCTCTCCCCGACCCTGCAGGCGACCCGCAGCAACTACACCGGC  3660
        N  I  N  A  V  Q  L  S  P  T  L  Q  A  T  R  S  N  Y  T  G

3661  GTCCACCGCGGCTCGAAGGTCCGGTTCATCGAGTACTTCAACGGCACGCGCCCGAGCCGG  3720
        V  H  R  G  S  K  V  R  F  I  E  Y  F  N  G  T  R  P  S  R

3721  ATCCTCGTCGACGTGCTCCAGTCCGAGCAGGGCGCGTGGTTCCTGCGCAAGCGCAACCGG  3780
        I  L  V  D  V  L  Q  S  E  Q  G  A  W  F  L  R  K  R  N  R
```

FIG. 4B–4

```
3781  AACATGGTCGTCGAGGTGTTCGACGACCTGCCCGAGCACCCGAACTTCCGGTGGCTGACC  3840
      N  M  V  V  E  V  F  D  D  L  P  E  H  P  N  F  R  W  L  T

3841  GTCGCGCAGCTGCGGGCGATGCTGCACCACGACAACGTGGTGAACATGGACCTGCGCACC  3900
      V  A  Q  L  R  A  M  L  H  H  D  N  V  V  N  M  D  L  R  T

3901  GTGCTGGCCTGCGTCCCGACCGCCGTGGAGCGGGACCGGGCCGACGACGTGCTCGCGCGC  3960
      V  L  A  C  V  P  T  A  V  E  R  D  R  A  D  D  V  L  A  R

3961  CTGCCCGAGGGCTCGTTCCAGGCCCGGCTGCTGCACTCGTTCATCGGCGCGGGCACCCCG  4020
      L  P  E  G  S  F  Q  A  R  L  L  H  S  F  I  G  A  G  T  P

4021  GCCAACAACATGAACAGCCTGCTGAGCTGGATCTCCGACGTGCGCGCCAGGCGCGAGTTC  4080
      A  N  N  M  N  S  L  L  S  W  I  S  D  V  R  A  R  R  E  F

4081  GTGCAGCGCGGCCGCCCGCTGCCCGACATCGAGCGCAGCGGGTGGATCCGCCGCGACGAC  4140
      V  Q  R  G  R  P  L  P  D  I  E  R  S  G  W  I  R  R  D  D

4141  GGCATCGAGCACGAGGAGAAGAAGTACTTCGACGTCTTCGGCGTCACGGTGGCGACCAGC  4200
      G  I  E  H  E  E  K  K  Y  F  D  V  F  G  V  T  V  A  T  S

4201  GACCGCGAGGTCAACTCGTGGATGCAGCCGCTGCTCTCGCCCGCCAACAACGGCCTGCTC  4260
      D  R  E  V  N  S  W  M  Q  P  L  L  S  P  A  N  N  G  L  L

4261  GCCCTGCTGGTCAAGGACATCGGCGGCACGTTGCACGCGCTCGTGCAGCTGCGCACCGAG  4320
      A  L  L  V  K  D  I  G  G  T  L  H  A  L  V  Q  L  R  T  E

4321  GCGGGCGGGATGGACGTCGCCGAGCTGGCGCCTACGGTGCACTGCCAGCCCGACAACTAC  4380
      A  G  G  M  D  V  A  E  L  A  P  T  V  H  C  Q  P  D  N  Y

4381  GCCGACGCGCCCGAGGAGTTCCGACCGGCCTATGTGGACTACGTGTTGAACGTGCCGCGC  4440
      A  D  A  P  E  E  F  R  P  A  Y  V  D  Y  V  L  N  V  P  R

4441  TCGCAGGTCCGCTACGACGCATGGCACTCCGAGGAGGGCGGCCGGTTCTACCGCAACGAG  4500
      S  Q  V  R  Y  D  A  W  H  S  E  E  G  G  R  F  Y  R  N  E

4501  AACCGGTACATGCTGATCGAGGTGCCCGCCGACTTCGACGCCAGTGCCGCTCCCGACCAC  4560
      N  R  Y  M  L  I  E  V  P  A  D  F  D  A  S  A  A  P  D  H

4561  CGGTGGATGACCTTCGACCAGATCACCTACCTGCTCGGGCACAGCCACTACGTCAACATC  4620
      R  W  M  T  F  D  Q  I  T  Y  L  L  G  H  S  H  Y  V  N  I

4621  CACGTGCGCAGCATCATCGCGTGCGCCTCGGCCGTCTACACCAGGACCGCCGGATGAAAC  4680
      H  V  R  S  I  I  A  C  A  S  A  V  Y  T  R  T  A  G  *
                                                        M  K  R

4681  GCGCGCTGACCGACCTGGCGATCTTCGGCGGCCCCGAGGCATTCCTGCACACCCTCTACG  4740
        A  L  T  D  L  A  I  F  G  G  P  E  A  F  L  H  T  L  Y  V
```

FIG. 4B-5

```
4741  TGGGCAGGCCGACCGTCGGGGACCGGGAGCGGTTCTTCGCCCGCCTGGAGTGGGCGCTGA  4800
        G  R  P  T  V  G  D  R  E  R  F  F  A  R  L  E  W  A  L  N

4801  ACAACAACTGGCTGACCAACGGCGGACCACTGGTGCGCGAGTTCGAGGGCCGGGTCGCCG  4860
        N  N  W  L  T  N  G  G  P  L  V  R  E  F  E  G  R  V  A  D

4861  ACCTGGCGGGTGTCCGCCACTGCGTGGCCACCTGCAACGCGACGGTCGCGCTGCAACTGG  4920
        L  A  G  V  R  H  C  V  A  T  C  N  A  T  V  A  L  Q  L  V

4921  TGCTGCGCGCGAGCGACGTGTCCGGCGAGGTCGTCATGCCTTCGATGACGTTCGCGGCCA  4980
        L  R  A  S  D  V  S  G  E  V  V  M  P  S  M  T  F  A  A  T

4981  CCGCGCACGCGGCGAGCTGGCTGGGGCTGGAACCGGTGTTCTGCGACGTGGACCCCGAGA  5040
        A  H  A  A  S  W  L  G  L  E  P  V  F  C  D  V  D  P  E  T

5041  CCGGCCTGCTCGACCCCGAGCACGTCGCGTCGCTGGTCACACCGCGGACGGGCGCGATCA  5100
        G  L  L  D  P  E  H  V  A  S  L  V  T  P  R  T  G  A  I  I

5101  TCGGCGTGCACCTCTGGGGCAGGCCCGCTCCGGTCGAGGCGCTGGAGAAGATCGCCGCCG  5160
        G  V  H  L  W  G  R  P  A  P  V  E  A  L  E  K  I  A  A  E

5161  AGCACCAGGTCAAACTCTTCTTCGACGCCGCGCACGCGCTGGGCTGCACCGCCGGCGGGC  5220
        H  Q  V  K  L  F  F  D  A  A  H  A  L  G  C  T  A  G  G  R

5221  GGCCGGTCGGCGCCTTCGGCAACGCCGAGGTGTTCAGCTTCCACGCCACGAAGGCGGTCA  5280
        P  V  G  A  F  G  N  A  E  V  F  S  F  H  A  T  K  A  V  T

5281  CCTCGTTCGAGGGCGGCGCCATCGTCACCGACGACGGGCTGCTGGCCGACCGCATCCGCG  5340
        S  F  E  G  G  A  I  V  T  D  D  G  L  L  A  D  R  I  R  A

5341  CCATGCACAACTTCGGGATCGCACCGGACAAGCTGGTGACCGATGTCGGCACCAACGGCA  5400
        M  H  N  F  G  I  A  P  D  K  L  V  T  D  V  G  T  N  G  K

5401  AGATGAGCGAGTGCGCCGCGGCGATGGGCCTCACCTCGCTCGACGCCTTCGCCGAGACCA  5460
        M  S  E  C  A  A  A  M  G  L  T  S  L  D  A  F  A  E  T  R

5461  GGGTGCACAACCGCCTCAACCACGCGCTCTACTCCGACGAGCTCCGCGACGTGCGCGGCA  5520
        V  H  N  R  L  N  H  A  L  Y  S  D  E  L  R  D  V  R  G  I

5521  TATCCGTGCACGCGTTCGATCCTGGCGAGCAGAACAACTACCAGTACGTGATCATCTCGG  5580
        S  V  H  A  F  D  P  G  E  Q  N  N  Y  Q  Y  V  I  I  S  V

5581  TGGACTCCGCGGCCACCGGCATCGACCGCGACCAGTTGCAGGCGATCCTGCGAGCGGAGA  5640
        D  S  A  A  T  G  I  D  R  D  Q  L  Q  A  I  L  R  A  E  K

5641  AGGTTGTGGCACAACCCTACTTCTCCCCCGGGTGCCACCAGATGCAGCCGTACCGGACCG  5700
        V  V  A  Q  P  Y  F  S  P  G  C  H  Q  M  Q  P  Y  R  T  E
```

FIG. 4B–6

```
5701  AGCCGCCGCTGCGGCTGGAGAACACCGAACAGCTCTCCGACCGGGTGCTCGCGCTGCCCA  5760
        P  P  L  R  L  E  N  T  E  Q  L  S  D  R  V  L  A  L  P  T

5761  CCGGCCCCGCGGTGTCCAGCGAGGACATCCGGCGGGTGTGCGACATCATCCGGCTCGCCG  5820
        G  P  A  V  S  S  E  D  I  R  R  V  C  D  I  I  R  L  A  A

5821  CCACCAGCGGCGAGCTGATCAACGCGCAATGGGACCAGAGGACGCGCAACGGTTCGTGAC  5880
        T  S  G  E  L  I  N  A  Q  W  D  Q  R  T  R  N  G  S  *

5881  GACCTGCGCCACAAGTGCCAGGAGGTTCGCTCCCCGATGAACACAACTCGTACGGCAACC  5940
                                          M  N  T  T  R  T  A  T

5941  GCCCAGGAAGCGGGGGTCGCCGACGCGGCGCGCCCGGACGTCGACCGGCGGGCGGTCGTG  6000
        A  Q  E  A  G  V  A  D  A  A  R  P  D  V  D  R  R  A  V  V

6001  CGGGCGCTGAGCTCGGAGGTCTCCCGCGTCACCGGCGCCGGTGACGGTGACGCCCACGTG  6060
        R  A  L  S  S  E  V  S  R  V  T  G  A  G  D  G  D  A  H  V

6061  CAGGCCGCCCGGCTCGCCGACCTCGCCGCGCACTACGGGGCGCACCCGTTCACGCCGCTG  6120
        Q  A  A  R  L  A  D  L  A  A  H  Y  G  A  H  P  F  T  P  L

6121  GAGCAGACGCGTGCGCGGCTCGGCCTGGACCGCGCGGAGTTCGCCCACCTGCTCGACCTG  6180
        E  Q  T  R  A  R  L  G  L  D  R  A  E  F  A  H  L  L  D  L

6181  TTCGGCCGCATCCCGGACCTGGGCACCGCGGTGGAGCACGGTCCGGCGGGCAAGTACTGG  6240
        F  G  R  I  P  D  L  G  T  A  V  E  H  G  P  A  G  K  Y  W

6241  TCCAACACGATCAAGCCGCTGGACGCCGCAGGCGCACTGGACGCGGCGGTCTACCGCAAG  6300
        S  N  T  I  K  P  L  D  A  A  G  A  L  D  A  A  V  Y  R  K

6301  CCTGCCTTCCCCTACAGCGTCGGCCTGTACCCCGGGCCGACGTGCATGTTCCGCTGCCAC  6360
        P  A  F  P  Y  S  V  G  L  Y  P  G  P  T  C  M  F  R  C  H

6361  TTCTGCGTGCGGGTGACCGGTGCCCGCTACGAGGCCGCATCGGTCCCGGCGGGCAACGAG  6420
        F  C  V  R  V  T  G  A  R  Y  E  A  A  S  V  P  A  G  N  E

6421  ACGCTGGCCGCGATCATCGACGAGGTGCCCACGGACAACCCGAAGGCGATGTACATGTCG  6480
        T  L  A  A  I  I  D  E  V  P  T  D  N  P  K  A  M  Y  M  S

6481  GGCGGGCTCGAGCCGCTGACCAACCCCGGTCTCGGCGAGCTGGTGTCGCACGCCGCCGGG  6540
        G  G  L  E  P  L  T  N  P  G  L  G  E  L  V  S  H  A  A  G

6541  CGCGGTTTCGACCTCACCGTCTACACCAACGCCTTCGCCCTCACCGAGCAGACGCTGAAC  6600
        R  G  F  D  L  T  V  Y  T  N  A  F  A  L  T  E  Q  T  L  N

6601  CGCCAGCCCGGCCTGTGGGAGCTGGGCGCGATCCGCACGTCCCTCTACGGGCTGAACAAC  6660
        R  Q  P  G  L  W  E  L  G  A  I  R  T  S  L  Y  G  L  N  N
```

FIG. 4B–7

```
6661  GACGAGTACGAGACGACCACCGGCAAGCGCGGCGCTTTCGAACGCGTCAAGAAGAACCTG  6720
      D  E  Y  E  T  T  T  G  K  R  G  A  F  E  R  V  K  K  N  L

6721  CAGGGCTTCCTGCGGATGCGCGCCGAGCGGGACGCGCCGATCCGGCTCGGCTTCAACCAC  6780
      Q  G  F  L  R  M  R  A  E  R  D  A  P  I  R  L  G  F  N  H

6781  ATCATCCTGCCGGGACGGGCCGACCGGCTCACCGACCTCGTCGACTTCATCGCCGAGCTC  6840
      I  I  L  P  G  R  A  D  R  L  T  D  L  V  D  F  I  A  E  L

6841  AACGAGTCCAGCCCGCAACGGCCGCTGGACTTCGTGACGGTGCGCGAGGACTACAGCGGC  6900
      N  E  S  S  P  Q  R  P  L  D  F  V  T  V  R  E  D  Y  S  G

6901  CGCGACGACGGCCGGCTGTCGGACTCCGAGCGCAACGAGCTGCGCGAGGGCCTGGTGCGG  6960
      R  D  D  G  R  L  S  D  S  E  R  N  E  L  R  E  G  L  V  R

6961  TTCGTCGACTACGCCGCCGAGCGGACCCCGGGCATGCACATCGACCTGGGCTACGCCCTG  7020
      F  V  D  Y  A  A  E  R  T  P  G  M  H  I  D  L  G  Y  A  L

7021  GAGAGCCTGCGGCGGGGTGTGGACGCCGAGCTGCTGCGCATCCGGCCGGAGACGATGCGT  7080
      E  S  L  R  R  G  V  D  A  E  L  L  R  I  R  P  E  T  M  R

7081  CCCACCGCGCACCCCCAGGTCGCGGTGCAGATCGACCTGCTCGGCGACGTCTACCTCTAC  7140
      P  T  A  H  P  Q  V  A  V  Q  I  D  L  L  G  D  V  Y  L  Y

7141  CGCGAGGCGGGCTTCCCGGAGCTGGAGGGCGCCACCCGCTACATCGCGGGCCGGGTCACC  7200
      R  E  A  G  F  P  E  L  E  G  A  T  R  Y  I  A  G  R  V  T

7201  CCGTCGACCAGCCTGCGCGAGGTGGTGGAGAACTTCGTGCTGGAGAACGAGGGCGTGCAG  7260
      P  S  T  S  L  R  E  V  V  E  N  F  V  L  E  N  E  G  V  Q

7261  CCCCGCCCCGGCGACGAGTACTTCCTCGACGGCTTCGACCAGTCGGTGACCGCACGGCTC  7320
      P  R  P  G  D  E  Y  F  L  D  G  F  D  Q  S  V  T  A  R  L

7321  AACCAGCTCGAACGAGACATCGCCGACGGGTGGGAGGACCACCGCGGCTTCCTGCGCGGA  7380
      N  Q  L  E  R  D  I  A  D  G  W  E  D  H  R  G  F  L  R  G

7381  AGGTGAACCGGAGTTGCGAGTACGTGAGCTGGCGGTGGCGGGCGGTTTCGAGTTCACCCC  7440
      R  *                                 V  A  G  G  F  E  F  T  P

7441  CGACCCGAAGCAGGACCGGCGGGGCCTGTTCGTGTCTCCGCTGCAGGACGAGGCGTTCGT  7500
       D  P  K  Q  D  R  R  G  L  F  V  S  P  L  Q  D  E  A  F  V

7501  GGGCGCGGTGGGCCATCGGTTCCCCGTCGCCCAGATGAACCACATCGTCTCCGCCCGGGG  7560
       G  A  V  G  H  R  F  P  V  A  Q  M  N  H  I  V  S  A  R  G

7561  CGTGCTGCGCGGGCTGCACTTCACCACCACCCCGCCGGGGCAGTGCAAGTACGTCTACTG  7620
       V  L  R  G  L  H  F  T  T  T  P  P  G  Q  C  K  Y  V  Y  C
```

FIG. 4B–8

```
7621  CGCGCGCGGCCGGGCGCTCGACGTCATCGTCGACATCCGGGTCGGCTCGCCGACGTTCGG   7680
        A  R  G  R  A  L  D  V  I  V  D  I  R  V  G  S  P  T  F  G

7681  GAAGTGGGACGCGGTGGAGATGGACACCGAGCACTTCCGGGCGGTCTACTTCCCCAGGGG   7740
        K  W  D  A  V  E  M  D  T  E  H  F  R  A  V  Y  F  P  R  G

7741  CACCGCGCACGCCTTCCTCGCGCTTGAGGACGACACCCTGATGTCGTACCTGGTCAGCAC   7800
        T  A  H  A  F  L  A  L  E  D  D  T  L  M  S  Y  L  V  S  T

7801  GCCGTACGTGGCCGAGTACGAGCAGGCGATCGACCCGTTCGACCCCGCGCTGGGTCTGCC   7860
        P  Y  V  A  E  Y  E  Q  A  I  D  P  F  D  P  A  L  G  L  P

7861  GTGGCCCGCGGACCTGGAGGTCGTGCTCTCCGACCGCGACACGGTGGCCGTGGACCTGGA   7920
        W  P  A  D  L  E  V  V  L  S  D  R  D  T  V  A  V  D  L  E

7921  GACCGCCAGGCGGCGAGGGATGCTGCCCGACTACGCCGACTGCCTCGGCGAGGAGCCCGC   7980
        T  A  R  R  R  G  M  L  P  D  Y  A  D  C  L  G  E  E  P  A

7981  CAGCACCGGCAGGTGACGGGTCCCGAGCACGATCTGTTCGAAGTGGCGCAGGCGCTCGTC   8040
        S  T  G  R  *

8041  GTCGCGGTCGA  8051
```

FIG. 4B–9

```
                1                                                  50
      eryBIV vngisdsprq litllGaSGf vGsavlreLr dhpv.rlrav sRggapavpp
        ascF .......mk. .llitGvSGy iGshlvnyLa nlggyeiygi sRneildqdi
        rfbJ ....mtflke yvivsGaSGf iGkhlleaLk k.sgisvvai tRdvi..knn
        strL ...mspyprp rwlvtGaSGm lGreltplLd rrga.avtal gRghl.ditd
   Consensus ---------- -----G-SG- -G-------L- ---------- -R--------

               51                                                 100
      eryBIV gaaevedlra dllepgraaa aiedadvivh lvahaAggst wrsatsdpe.
        ascF nql.llniki fqldrdslpd ilkrvrpdv. .vihlAscfl sqhsyknike
        rfbJ sna.lanvrw cswdniellv eelsidsali giihlAtey. .ghktsslin
        strL gaa..... .....vrsav aehrpavvvn caawtAvd.. ..eaesepal
   Consensus ---------- ---------- ---------- -----A---- ----------

              101                                                 150
      eryBIV aervNvglmh dLvgalhdrr rstppvllya staqaanpsa asryaqqkte
        ascF iiksNvefpt eLlea..... ......mndv gvkkiintgt swqcfnsdty
        rfbJ iedaNvikpl kLldl..... .......aiky radiflntd. sffakkdfny
        strL amavNgegpr hLaqa....c ravgavllql stdyvfpgsg grpyredhpt
   Consensus ----N----- -L-------- ---------- ---------- ----------
```

FIG. 5A-1

```
           151                                                      200
   eryBIV  aerilrkatd egrvrgvilr lpa....... ..vygqsGps gpmgrgvvaa
     ascF  npvnlyaask qafedilkfy inaegfsain lklfdtyGgv dkrrklis.l
     rfbJ  qhmrpyiitk rhfdeighyy anmhdisfvn mrlehvyGpg dgenkfipyi
     strL  gprtvygctk rageravlev lpdtgyivrt awlygagGp. .....nfvak
Consensus  ---------- ---------- ---------- -------G-- ----------

           201                                                      250
   eryBIV  mirralagep ltmwhdggvr rdl.lhveDv atAFaaaleh hdalag...g
     ascF  lddiaknnkq ldmspgeqll d..lvhinDv crAFkiaidk lcelpseyvv
     rfbJ  idclnkkqsc vkcttgeqir d..fifvdDv vnAYltilen rkevps..yt
     strL  mirleadedt vlvvddqhgq ptwtadlaDr laALgaaala gtapagiyha
Consensus  ---------- ---------- --------D- --AF------ ----------
```

FIG. 5A-2

```
                251                                                  300
   eryBIV   twalgadrse plgDifravs gsva..rqtg spavdvvtvp apehaeandf
     ascF   sygvsnkyrv tlkElvsiye ..rv...nnv klninfgtre yrnrevmvpc
     rfbJ   eyqvgtgagv slkDflvylq ntmm...pgs ssifefgaie qrdneimfsv
     strL   tntggttwna lapEtfrllg adparvrptt slalarpavr .prysvldqs
Consensus   ---------- ---------- ---------- ---------- ----------

                301                                     345
   eryBIV   rsddidstef rsrtgwrprv sltdgidrtv aaltpteeh .....
     ascF   .tniqnl... ...pdwevvi plsqglky.. .......... .....
     rfbJ   .annknlkam gwkpnfdykk gieellkrl. .......... .....
     strL   rwkaaglepl rhw.....ra altesf.... palcgragrp vpgpr
Consensus   ---------- ---------- ---------- ---------- -----
```

FIG. 5A-3

```
             1                                                    50
  eryBVII .... ...... ......Vagg feftpdpkqD rRGlFvspLq deaFvgavGh
     strM.           .. .vrplsVqga wlsetrafaD dRGeFqelYs arsLrgalGy
     rfbC  .........vm ivikta Ipdv lilepkvfgD eRGfFfesYn qqtFeeliGr
     rfbF  .........m kctklsIpev ilfeprifeD dRGhFfesFn lakFqesiGr
     ascE  lgvivphylm ifkkldIegc yliefnkfiD sRGtFvktFh sdfFsen.Gi
Consensus  ---------- ------I--- ---------D -RG-F----- ---F----G-

             51                                                  100
  eryBVII  rfpvaqmnhi vSargVlRGl HFtttppgqc KyVycarGra LDVivDiRvg
     strM  dpgvaqvnrs vSrrgVlRGv HFaqlppsqa KyVtclsGav LDVvvDiRtg
     rfbC  kvtfvqdnhs kSkknVlRGl HFqrgenaqg KlVrcavGev FDVavDiRke
     rfbF  qvtfvqsnes ySkqnViRGl HYq.virpqg KlVrvveGev FDIavDlRks
     ascE  vldmreefys iSaknViRGm HFqmppaehd KlVycvnGav LDVilDiRkd
Consensus  ---------- -S---V-RG- HF-------- K-V----G-- LDV--D-R--

             101                                                 150
  eryBVII  SpTFgkWdaV eMdtehfrav YfprGtaHaF laLeddtlms Ylvstpyvae
     strM  SpTYraWeaV rLddph.rsl YveaGlgHsF maLtddavvv Yltsqgyaag
     rfbC  SpTFgqWvgV nLsaenkrql WipeGfaHgF vtLseyaefl Ykatnyysps
     rfbF  SpTFgqWvgV lLsdknnhql WipeGfgHgF qvLspsakfq Ymvtdywype
     ascE  SkTYgeYfsI eLsyenslal WvpkGlaHgF lsLadnsimf Yktssvhnve
Consensus  S-TF--W--V -L-------- W---G--H-F --L------- Y---------
```

FIG. 5B-1

```
            151                                                         200
eryBVII     yEqaIdpfdp alglpwpadl evvlsDrdtv avdletarrr gmlpdyadcl
strM        rEhgVhpldp dlgiawpdgi epvlsEkdrq apgiaemerr gllpdyeecl
rfbC        sEgsIlwnde aigiewp... fsqlpE.... ...lsakdaa aplldqallt
rfbF        hDrcIrfnds dinikwk... egiisEqqvi eyklsskdis gnsladaevf
ascE        cDsgIkwnsf gfkwpid... npiisEkdns lcyfdefdss f.........
Consensus   -E--I----- ---------- -----E---- ---------- ----------

            201        212
eryBVII     geepastgr ..
strM        afrrslcerg tg
rfbC        e......... ..
rfbF        .......... ..
ascE        .......... ..
Consensus   ---------- --
```

FIG. 5B-2

FIG. 5C-1

```
                1                                                   50
eryCIV  ........mk raltdLaifg gpeaFlhtly vgrptVgd.. ....rerFfa
eryCI   ........md vpfldLqa.. ayleLrsdid qAcrrVlg.. ....sgwY..
ascC    ...msqeelr qqiaeLvaq. aetaMapkpf eAgksVvpps gkvigtkelq
dnrJ    .......... vstyvWqyln eyreeradil dAvetVfe.. ....sgqL..
prg1    .......... .......... .......... .......... ..........
strS    .......... ..mssFqelp rwpqLtdddi eAavaalr.. ....snrL..
Consensus ---------- -----L---- ----L----- -A---V---- ----------

                51                                                  100
eryCIV  rlewalnnnw Ltnggplvre FEgrvAdL.. aGvrHcVatc natvAlqLvL
eryCI   .......... Lhgpe..nea FEaeFAaY.. cenaHcVtvg SGcdAleLsL
ascC    lmveasldgw L.ttgrfnda FEkkLgeYl. .Gvpyvlttt SGssAnlLAL
dnrJ    .......... ilgts..vrs FEeeFAaY.. hGlpyctgvd nGtnAlvLgL
prg1    .......... ..sgp..igq LEaeFlaFld hGvryaVtfn SGtsAllaAy
strS    .......... vgqgnstvee FEaaLAa..g qGveHaVavs tGtaAvhLAL
Consensus ---------- L--------- FE---A-Y-- -G--H-V--- SG--A--LAL

                101                                                 150
eryCIV  rAs.......  .dV..sgEVV mPsmTFaaTa haaswlGleP VFcDVDpeTg
eryCI   vAL.......  .gVgqGDEVI vPshTFiaTw lgV.pvGAvP VpVEpEgvsh
ascC    tALtspklgv ralkPGDEVI tvaagFptTv nptiqnGliP VFVDVDipT.
dnrJ    rAL.......  .gIgPGDEVV tvsnTaapTv vaIdavGAtP VFVDVhe...
prg1    fAL.......  .gVreGvEaa gPalTYhaal spVfalrgdv VLVDIDpvsr
strS    hAL.......  .dVgPGDEVI vPthTFigsa spVtylGArP VFaDVtpdTh
Consensus -AL------- --V-PGDEVI -P--TF--T- --V---GA-P VFVDVD--T-
```

FIG. 5C-2

```
           151                                                200
eryCIV     ......lLDP ehVaslVTpr TgAIigVHlW GrpapvEale kIAaeHqvkl
eryCI      t......LDP alVEqAITpr TaAIlpVHlY GhpaDLDalr aIAdrHgLal
ascC       .....ynvna slIEAAVsDk TkAImiaHtl GnlfDLaeVr rVAdkynLwl
dnrJ       ...enylMDt grlrsvIgpr TrcllpVHlY GqsvDMtpVl elAaeHdLkV
prgl       g......LDP kalEAAITEn TrvVtvVHqW GhpcDMDaIl gVAerygLrV
strS       c......LDP dsVkslIgEr TkAIvvVHin GiaaDMaalt aVAaeagvpV
Consensus  -------LDP --VEAAITE- T-AI--VH-Y G---DMD-V- -IA--H-L-V

           201                                                250
eryCIV     ffDaAhAlGc tagGrpVGaF GnaevFS.Fh atKavtsf.E GGAIVTdDglL
eryCI      VEDvAQAvGa rhrGhrVGag snaaaFS.FY PgKnlgAlGD GGAVVTtDpaL
ascC       IEDccdAlGs tYdGkmaGtF GdigtvS.FY Pahhitm.GE GGAVfTqsaeL
dnrJ       lEDcAQAhGa rrhGrlVGtq GhaaaFS.FY PtKvlgAyGD GGAVVTpDaev
prgl       lEDcshAhGs rYkGkvpGtF GdaavFS.Lq anKavyA.GE GGilVTdDalv
strS       IEDaAQAlGt eigGrpIGgF GdlacvSlFF eqKvitsgGE GGAVITdnpvy
Consensus  IED-AQA-G- -Y-G--VG-F G----FS-FY P-K---A-GE GGAVVT-D--L

           251                                                300
eryCIV     aDRiRaMhn FGiapdk... .......... .......... ....lvtdvG
eryCI      aERiRlLRn YG.....sk. .......... .......... qk.YvhevrG
ascC       ksiiesFRd WGrdcycapg cdntckkrfg qqlgslpfgy dhkYtyshlG
dnrJ       drRlRrLRy YG.....mg. .......... .......... eryYvvdtpG
prgl       qDRatlLat tG.....t.. .......... .......... .......vpG
strS       aERvRrLRs hGegpvsgs. .......... .......... pgmiwahevG
Consensus  -ER-R-LR- YG-------- ---------- ---------- ---Y-----G
```

```
            301                                                    350
eryCIV      tNgkMsEcaA AmgltsLdaF aetrvhnrln halysdeLrd vrGisvhafd
eryCI       tNaRLdELQA AvlrvkLrhL DdWnarRttl aqhyqteLkd vpGitlpeth
ascC        yNikitDMQA AcglaqLepi EeFvekRkan fkylkdaLqs cadf.ielpe
dnrJ        hNsRLdEvQA eilrrkLrrL DaYvegRrav arryeegLgd ldGlvlpt..
prg1        lghRr..... .......... .......... .......... ..........
strS        yNvRLtsvQA psaspsnkrL gdLveaRrrn aaylserLag veGlelpvep
Consensus   -N-RL-E-QA A-----L--L D-W---R--- -------L-- --G-------

            351                                                    400
eryCIV      pgeqnnyqYv iisVdsaatg idr..dqLqa iLraekVvaq pyFspgcHqm
eryCI       pwads..aWh lFvlrcenrD ......hLqr hLtdagVqtl ihYptpvHls
ascC        atensdpsWf gFpI.tlkeD sgvsridLvk fLdeakVgtr llFagnltrq
dnrJ        iaegndhvYy vYvVrhperD ......rile aLtaydIhln isYpwpvHtm
prg1        .......... .......... .......... .......... ..........
strS        p..gtthaYw kYaVrvvpgD grrsaadiaa hLrsrgVpvl lrYpyplHkq
Consensus   --------W- -Y-------D -------L-- -L----V--- --Y----H--
```

FIG. 5C-3

```
                                                                  450
          401
eryCIV    qpYrte...p plrLentEqL sdrvLaLPtg PavssEdirr Vcdiirlaat
eryCI     paYAdlgl.p pGsFpvaEsL agevLsLPig PhLsrEaaDh VIatlkaga
ascC      pyFhdvkyRv vGeLtntDRi mnqtFwigiy PgLthDhlDy VVskfeeffg
dnrJ      sgFAhlg.Yg pGdLpvtERL ageiFsLPmy PsLrpDaqEk VIdavrevvg
prgl      .......... .......... .......... .......... ..........
strS      paFAe...Yh gvsLpvaERL sqelLaLPsh PgLveghlDh aVeevrkava
Consensus --FA----Y- -G-L---ERL ----L-LP-- P-L--E--D- VV----------

          451       468
eryCIV    sgelinaqwd qrtrngs
eryCI     .......... .......
ascC      lnf....... .......
dnrJ      sl........ .......
prgl      .......... .......
strS      s......... .......
Consensus ---------- -------
```

FIG. 5C-4

```
                1                                                  50
      eryBV     vrVl1tsfAh rtHFqGlVPL AWALRtAGHD VRVaaqPALT DaVigAGLTA
    eryCIII     mrVvfssmAs ksHLfGlVPL AWAFRaAGHE VRVvasPALT EdItaAGLTA
       dnrS     mkVlvtafAm daHFnGvVPL AWALRaAGHD VRVasqPALT DsItrAGLTA
  Consensus     --V-----A- --HF-G-VPL AWALR-AGHD VRV---PALT D-I--AGLTA

               51                                                 100
      eryBV     VPVGsDhrlf divpevaaqV hrysfylDFy hreqelhswE fLlgmqeats
    eryCIII     VPVGtDvdlv dfmthaghdI idyvrslDFs erdpatltwE hLrgmqtvlt
       dnrS     VPVGtDhqvq aamgamapgV falhlnpDYl enrpelldlE fLeastsmlt
  Consensus     VPVG-D----- ----------V ---------DF- ----------E -L---------

              101                                                 150
      eryBV     rwvYpvvmnD sFVaeLVdFa rdWRPDLVlW EPFTFagaVA aracGaahAR
    eryCIII     ptfYalmspD tLIegMVsFc rkWRPDLViW EPLTFaapIA gavtGtphAR
       dnrS     aafYaqinnD sMIdeMVdFa awWRPDLVvW EPFTFggaVA aqvtGaaqAR
  Consensus     ---Y------D --I--MV-F- --WRPDLV-W EPFTF---VA ----G---AR
```

FIG. 5D-1

```
                151                                                   200
     eryBV      LLWGsDltgy frgrFqaqrl rrPpEdRpdp LgtWLTevag rfGv.eFgeD
   eryCIII      LLWGpDittr arqnFlgllp dqPeEhRegp LaeWLTwtle kyGgpaFdeE
      dnrS      LLWGpDlflr vhdrFqqvlh evPaErRdda LeeWLTwtle rhGa.aFgpE
 Consensus      LLWG-D---- ----F----- --P-E-R--- L--WLT---- --G---F--E

                201                                                   250
     eryBV      lavGqWsVDq lPpsfRLdTg meTVvaRtlp YNG..asVVP dWLkkgsatr
   eryCIII      vvvGqWtIDp aPaaiRLdTg lkTVgmRyvd YNG..psVVP eWLhdeperr
      dnrS      visGhWtIDq mPpsvRFaTa rpTVpmRfvp YNGpvpaVVP pWLradpgrp
 Consensus      ---G-W-ID- -P---RL-T- --TV--R--- YNG----VVP -WL-------

                251                                                   300
     eryBV      RIciTgGfsg lgla.adadq fartLaqlar fDgEIVvTgs gpdtsav...
   eryCIII      RVclTlGiss rensigqvs. ieelLgavgd vDaEIIaTfd aqqlegvani
      dnrS      RVllTqGite rstgftglpr agelLasiae lDaEVVaTvk aeereglppl
 Consensus      RV--T-G--- ---------- ----L----- -D-EIV-T-- ----------
```

FIG. 5D-2

```
                301                                                350
      eryBV  PdNIRlVdfv pMgvlLqnCA AiIHHGGaGt WaTAlhHGIP QisvaheWDc
   eryCIII  PhNVRtVgfv pMhalLptCA AtVHHGGpGs WhTAaiHGVP Qvilpdgwdt
      dnrS  PgNVRvVdsl sLhvvLpsCA AvVHHGGaGt WaTAalHGVP Qlalawqwdd
 Consensus  P-NVR-V---  -M---L--CA A-VHHGG-G-  W-TA--HGVP Q------WD-

                351                                                400
      eryBV  mlRgqqtaeL GAGIyLrp.. devdadslas altqVveDPt ytenAvklRe
   eryCIII  gvRaqrtqeF GAGIaLp..v peltpdqlre svkrVldDPa hragAarmRd
      dnrS  vfRagqlekL GAGIfLpphg egasagrvrd rlaqVlaEPs frqgAariRa
 Consensus  --R------L GAGI-L---- ---------- ----V--DP- ----A---R-

                401                433
      eryBV  EaLsdPtPqe IVprlEeLtr rhag..... ...
   eryCIII  DmLaePsPae VVgicEeLaa grrepr.... ...
      dnrS  EmLrtPaPga VVptlEqLta rhrapagggv rh
 Consensus  E-L--P-P-- VV---E-L-- ---------- ---
```

FIG. 5D-3

```
                1                                                    50
eryCVI     MYegg.fAel YDrfyrgRgK DYaaeaaqva rlvrdrlpsA ssLLDVACGT
  srmX     MYendsaAev YDllyqdR.K DYageaarvt dlirertpdA asLLDIACGT
  rdmD     MYgad.lArv YDlvhreRgK DYrardrggr rrgpaeqagA grLLDVACGT
Consensus  MY-----A-- YD-----R-K DY-------- ---------A --LLDVACGT

                51                                                  100
eryCVI     GtHLrrFAdL FddVtGlELS aaMievArpq LgGIpvlqGD MRdFaLdreF
  srmX     GtHLeaFAkL YdrVsGlELS ewMaarAeer LpGVtlhrGD MRaFdLgetF
  rdmD     GgHLrhFAdL FahVeGvELS epMaeeAraa LpGVtvhaGD MRdFrLgttF
Consensus  G-HL--FA-L F--V-G-ELS --M---A--- L-GV----GD MR-F-L---F

                101                                                 150
eryCVI     DaVtCMFsSI GhMrdgAELd qAlasfARHL apgGVvvVEP WWFpEdFlDG
  srmX     DaVvCMFsSI GyLettADLe dAvaamARHL tadGVlaVEP WYFpDtFlDG
  rdmD     DvVtCMFgSV GyMtsvAELg rAlrmfARHL epgGVavVDP WWFyEtFaDG
Consensus  D-V-CMF-SI G-M---AEL- -A----ARHL ---GV--VEP WWF-E-F-DG
```

FIG. 5E-1

```
                                                                200
      151
eryCVI yVagdvvrd. .gdltIsRVS HSvRaGgatr MEIHWvVAda vnGprHhvEh
  srmX hVsthalrta pgdqgVaRVS HStReGgrtr MEIHYlIAht aeGirHrsEv
  rdmD hVsadivtv. .dgvtVsRVS HSaRrGrtsh MDVHFvVAep gaGaqHfvDt
Consensus -V--------- -----V-RVS HS-R-G---- MEIH--VA-- --G--H--E-

                                                          241
      201
eryCVI yeitLFeRqq YEkAFtaAGc aVqYleggps grGLFvGvRg
  srmX dyltLFsRae YEaAYrkAGl dVeYvvtgeg spGFFlGtRr
  rdmD hiisLFsRse YEqAFrdAGf aVeYlpeaps grGLFvGvRg
Consensus ----LF-R-- YE-AF--AG- -V-Y------ --GLF-G-R- -
```

FIG. 5E-2

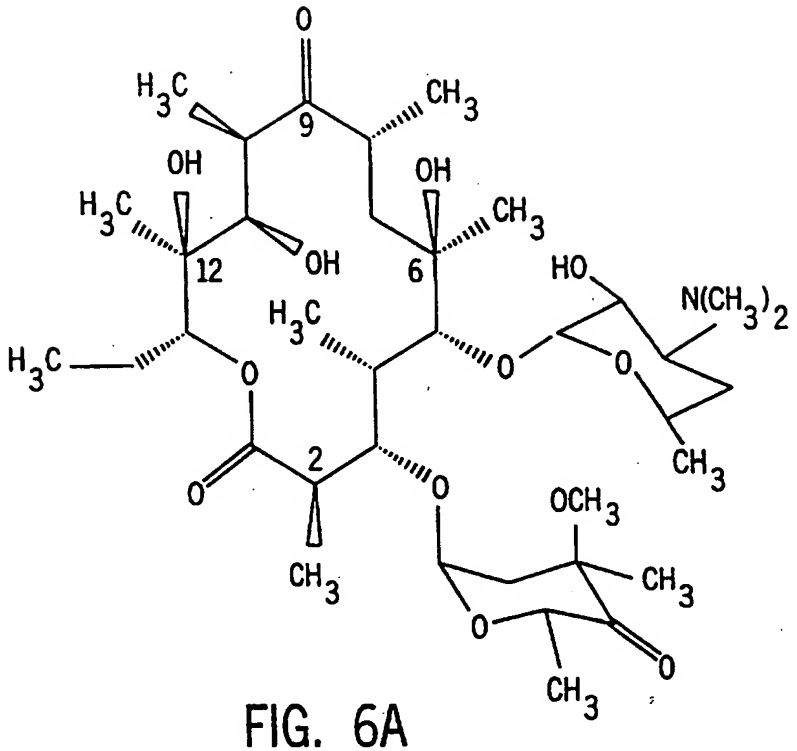

FIG. 6A

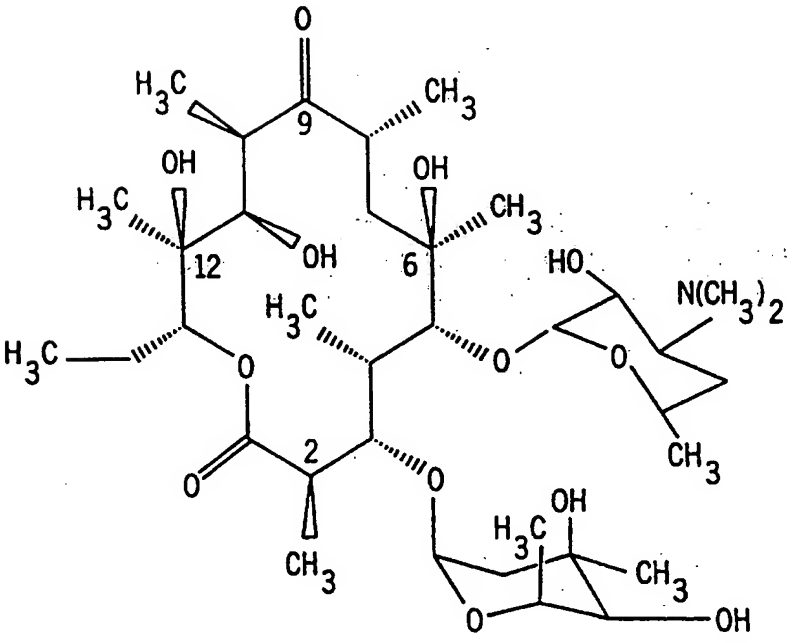

FIG. 6B

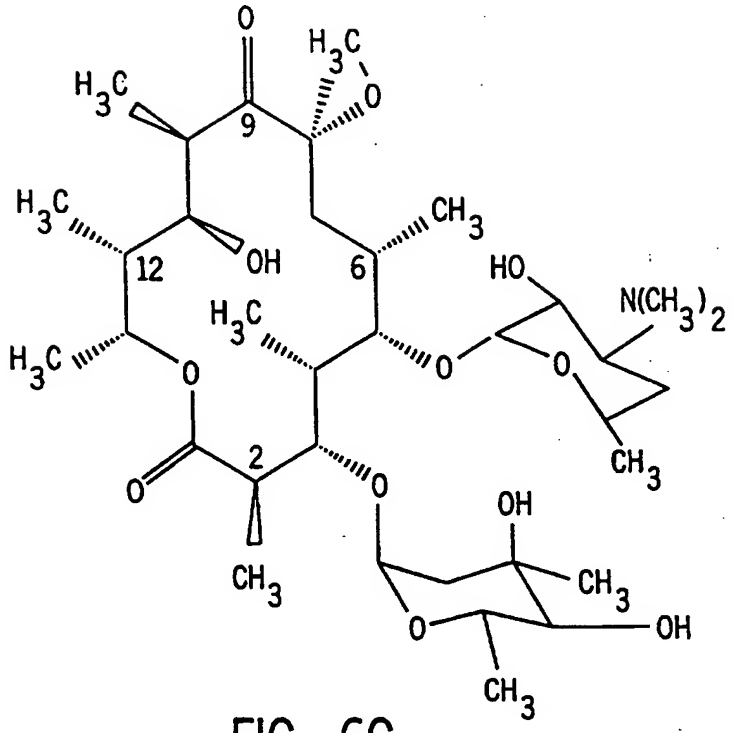

FIG. 6C

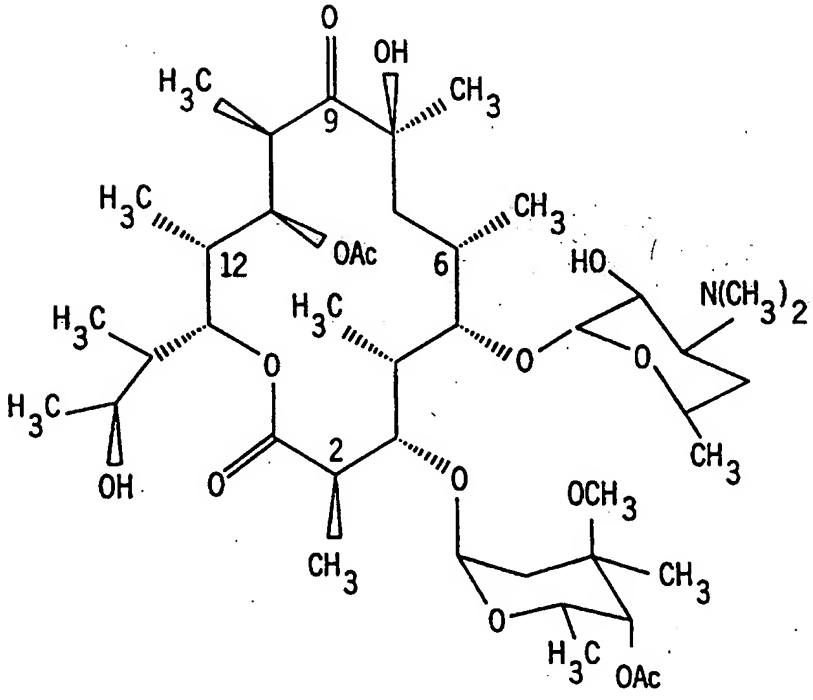

FIG. 6D

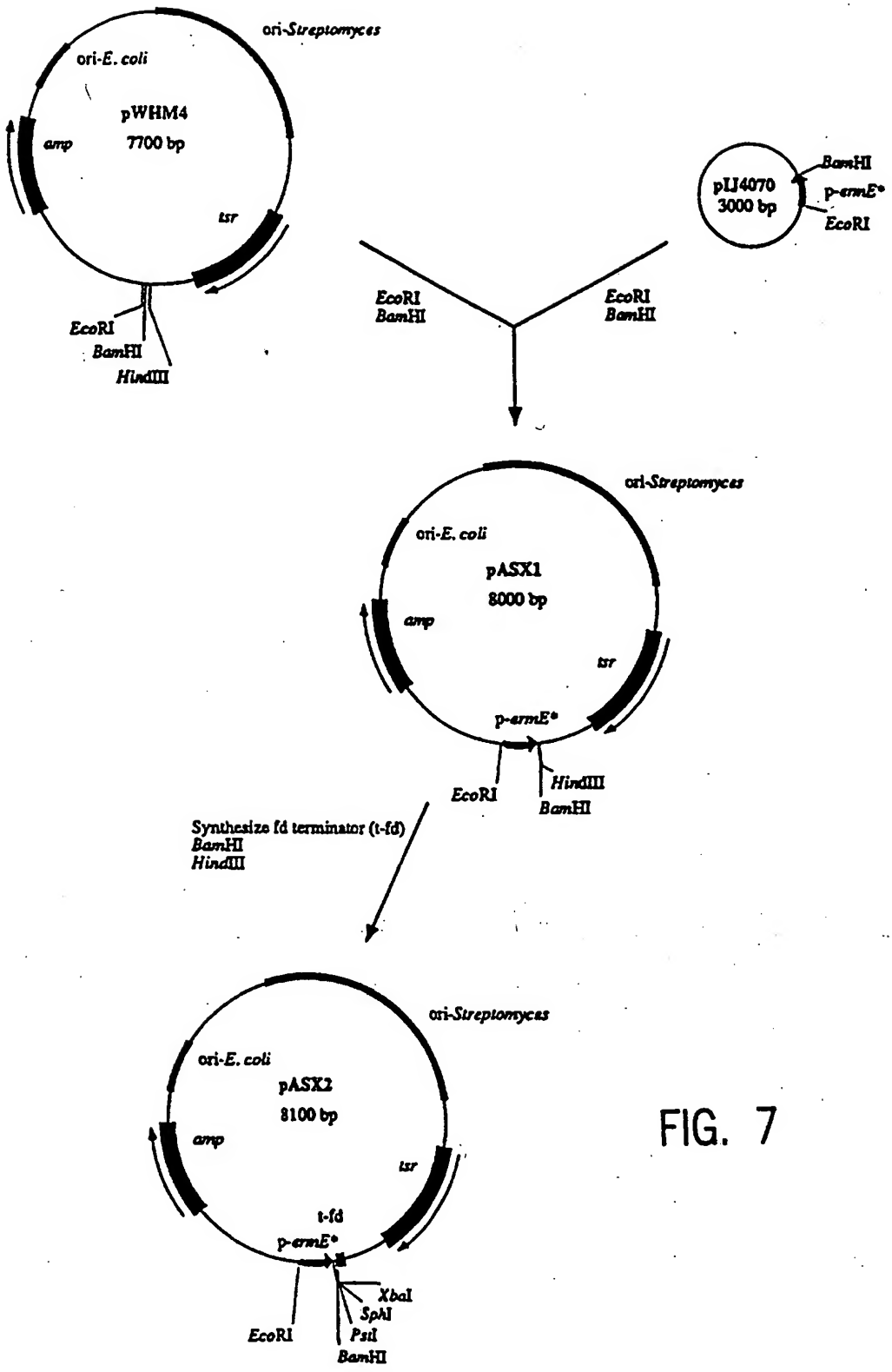
ori-Streptomyces
ori-E. coli
pWHM4
7700 bp
amp
tsr
EcoRI
BamHI
HindIII
pLJ4070
3000 bp
BamHI
p-ermE*
EcoRI
EcoRI
BamHI
EcoRI
BamHI
ori-Streptomyces
ori-E. coli
pASX1
8000 bp
amp
tsr
p-ermE*
EcoRI
HindIII
BamHI
Synthesize fd terminator (t-fd)
BamHI
HindIII
ori-Streptomyces
ori-E. coli
pASX2
8100 bp
amp
tsr
t-fd
p-ermE*
EcoRI
XbaI
SphI
PstI
BamHI

FIG. 7

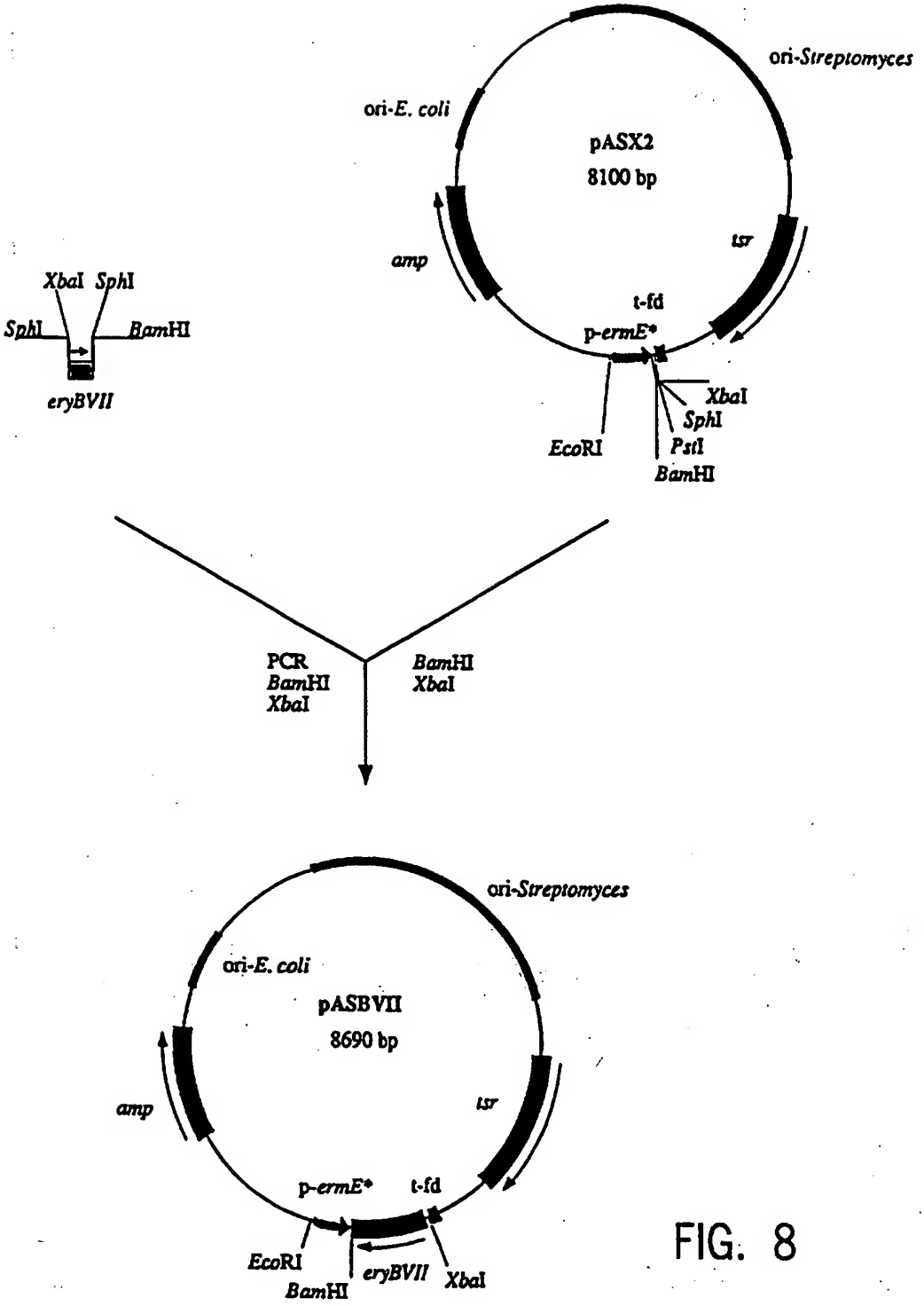
ori-Streptomyces
ori-E. coli
pASX2
8100 bp
tsr
amp
t-fd
p-ermE*
XbaI
SphI
PstI
BamHI
EcoRI
XbaI SphI
SphI
BamHI
eryBVII
PCR
BamHI
XbaI
BamHI
XbaI
ori-Streptomyces
ori-E. coli
pASBVII
8690 bp
amp
tsr
p-ermE*
t-fd
EcoRI
BamHI
eryBVII
XbaI

FIG. 8

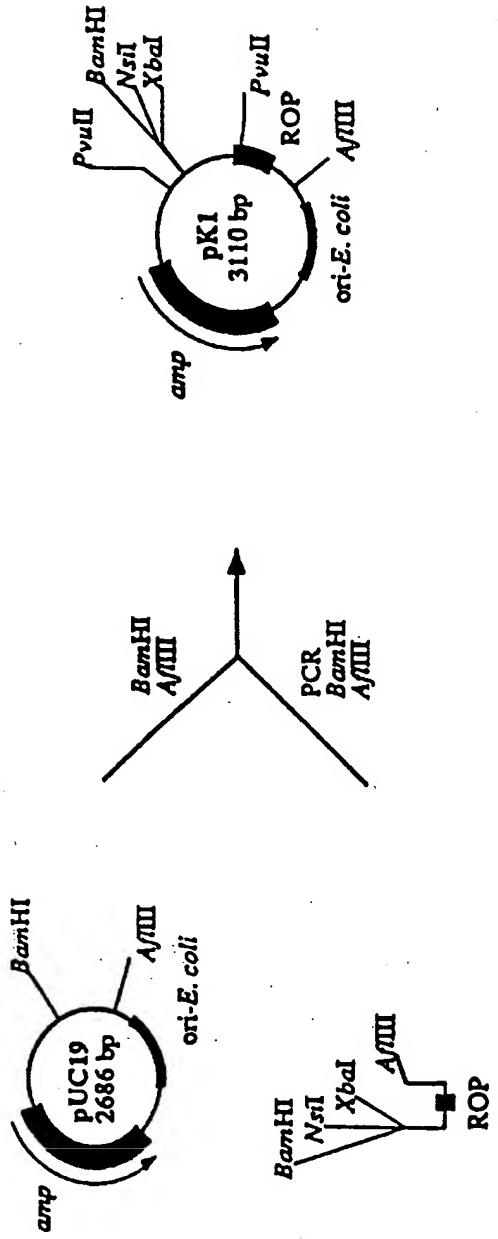
BamHI
NsiI
XbaI
PvuII
PvuII
ROP
AflIII
pK1
3110 bp
ori-E. coli
amp
BamHI
AflIII
PCR
BamHI
AflIII
BamHI
AflIII
pUC19
2686 bp
ori-E. coli
amp
AflIII
BamHI
NsiI
XbaI
ROP

FIG. 9A

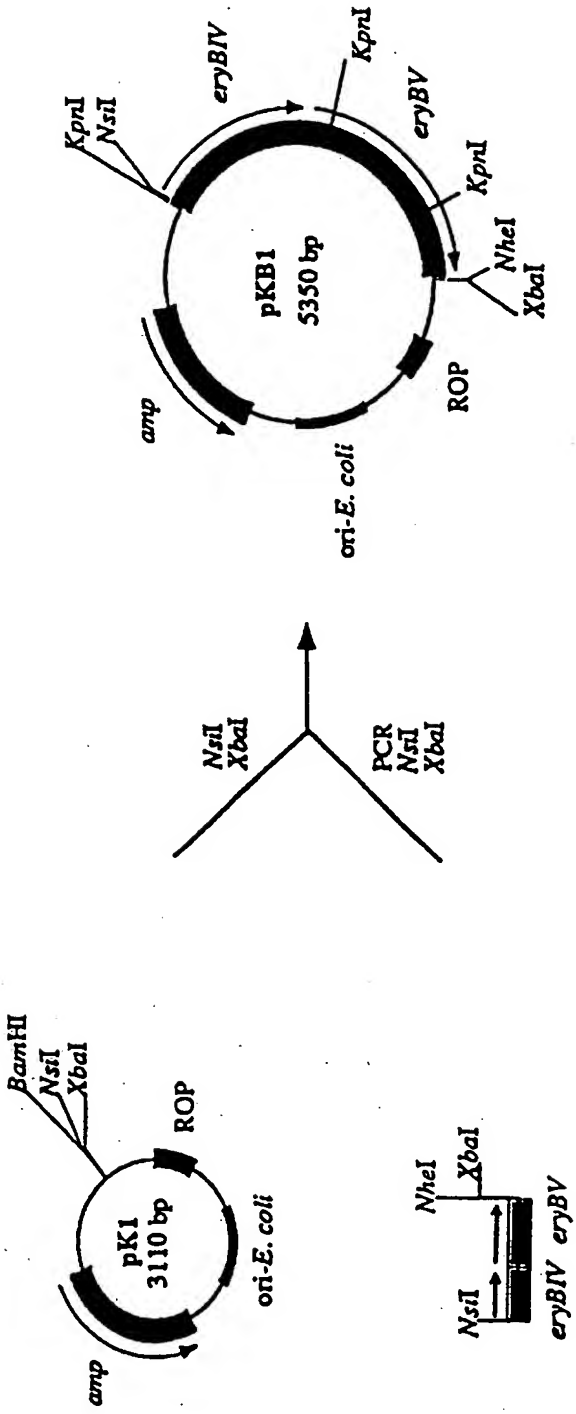
KpnI
NsiI
eryBIV
KpnI
eryBV
KpnI
NheI
XbaI
pKB1
5350 bp
ROP
amp
ori-E. coli
NsiI
XbaI
PCR
NsiI
XbaI
BamHI
NsiI
XbaI
ROP
pK1
3110 bp
ori-E. coli
amp
NheI
XbaI
NsiI
eryBIV eryBV

FIG. 9B

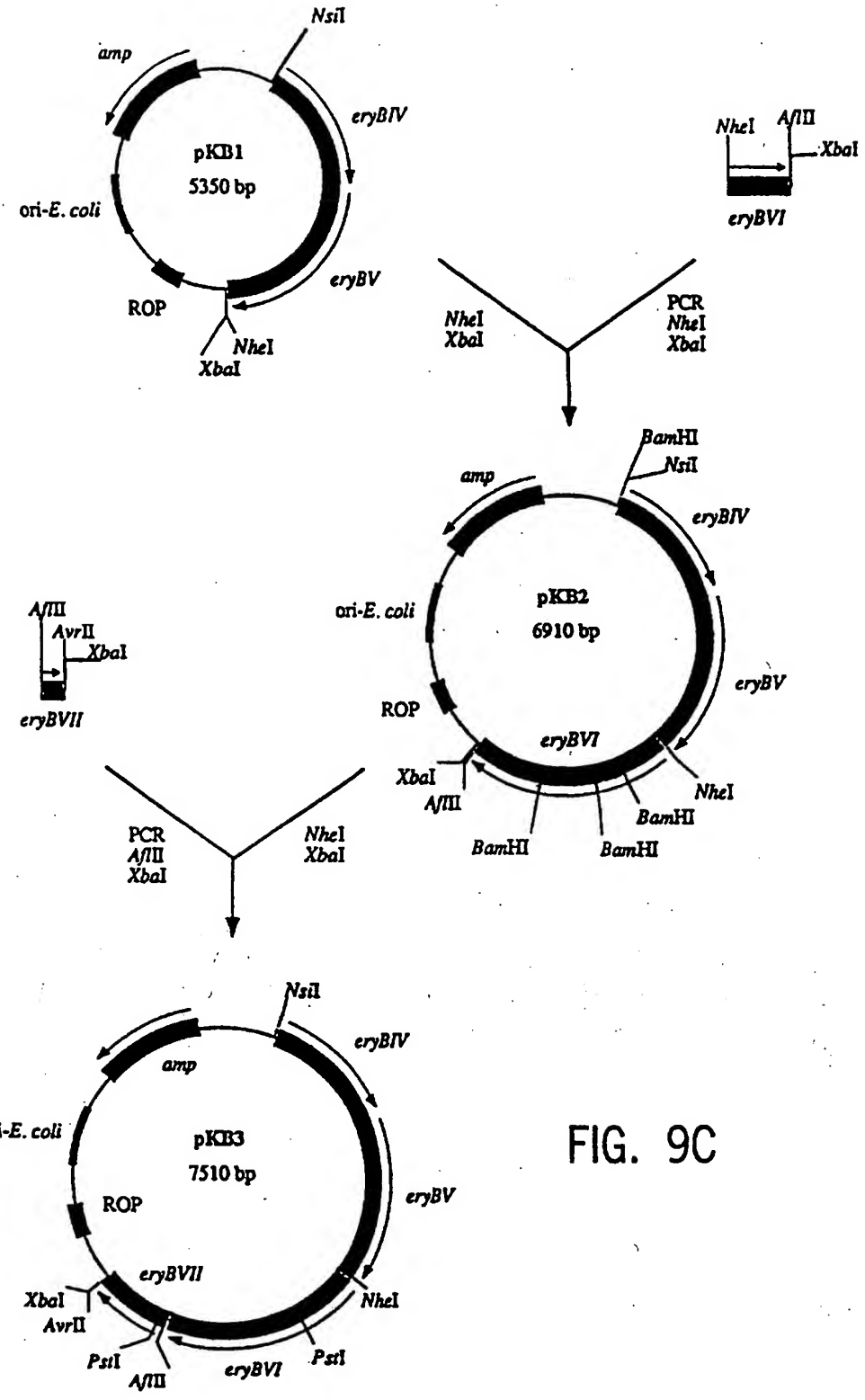
NsiI
amp
eryBIV
pKB1
5350 bp
ori-E. coli
eryBV
ROP
NheI
XbaI
NheI
AflII
XbaI
eryBVI
PCR
NheI
XbaI
NheI
XbaI
BamHI
NsiI
amp
eryBIV
pKB2
6910 bp
ori-E. coli
eryBV
ROP
eryBVI
XbaI
AflII
NheI
BamHI
BamHI
BamHI
AflII
AvrII
XbaI
eryBVII
PCR
AflII
XbaI
NheI
XbaI
NsiI
eryBIV
amp
pKB3
7510 bp
ori-E. coli
eryBV
ROP
eryBVII
XbaI
AvrII
NheI
PstI
AflII
eryBVI
PstI

FIG. 9C

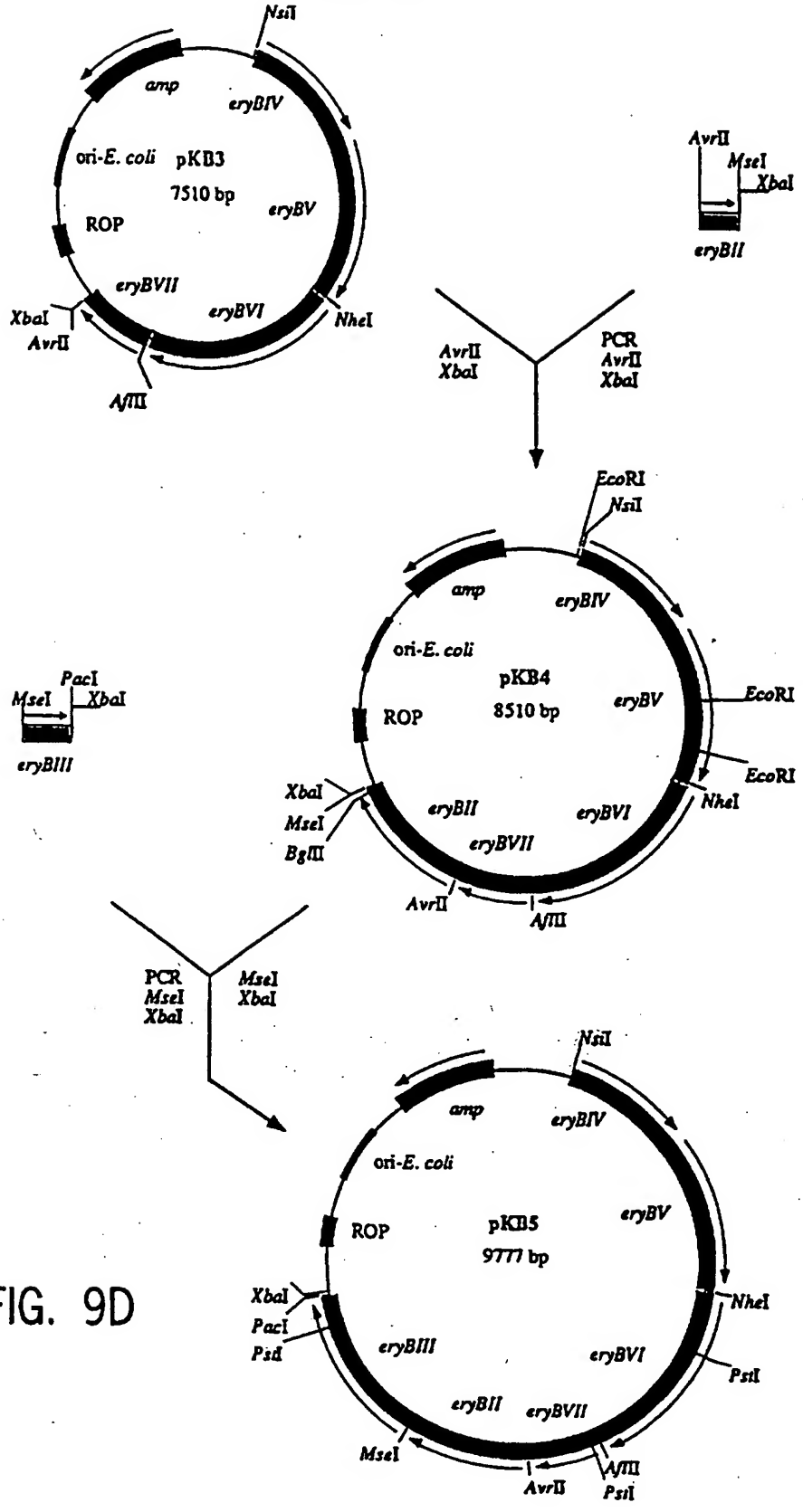
NsiI
amp
eryBIV
ori-E. coli
pKB3
7510 bp
eryBV
ROP
eryBVII
eryBVI
XbaI
AvrII
NheI
AflII
AvrII
MseI
XbaI
eryBII
AvrII
XbaI
PCR
AvrII
XbaI
EcoRI
NsiI
amp
eryBIV
ori-E. coli
pKB4
8510 bp
ROP
eryBV
EcoRI
EcoRI
NheI
XbaI
MseI
BglII
eryBII
eryBVI
eryBVII
AvrII
AflII
PacI
MseI
XbaI
eryBIII
PCR
MseI
XbaI
MseI
XbaI
NsiI
amp
eryBIV
ori-E. coli
ROP
pKB5
9771 bp
eryBV
NheI
XbaI
PacI
PstI
eryBIII
eryBVI
PstI
eryBII
eryBVII
MseI
AflII
AvrII
PstI

FIG. 9D

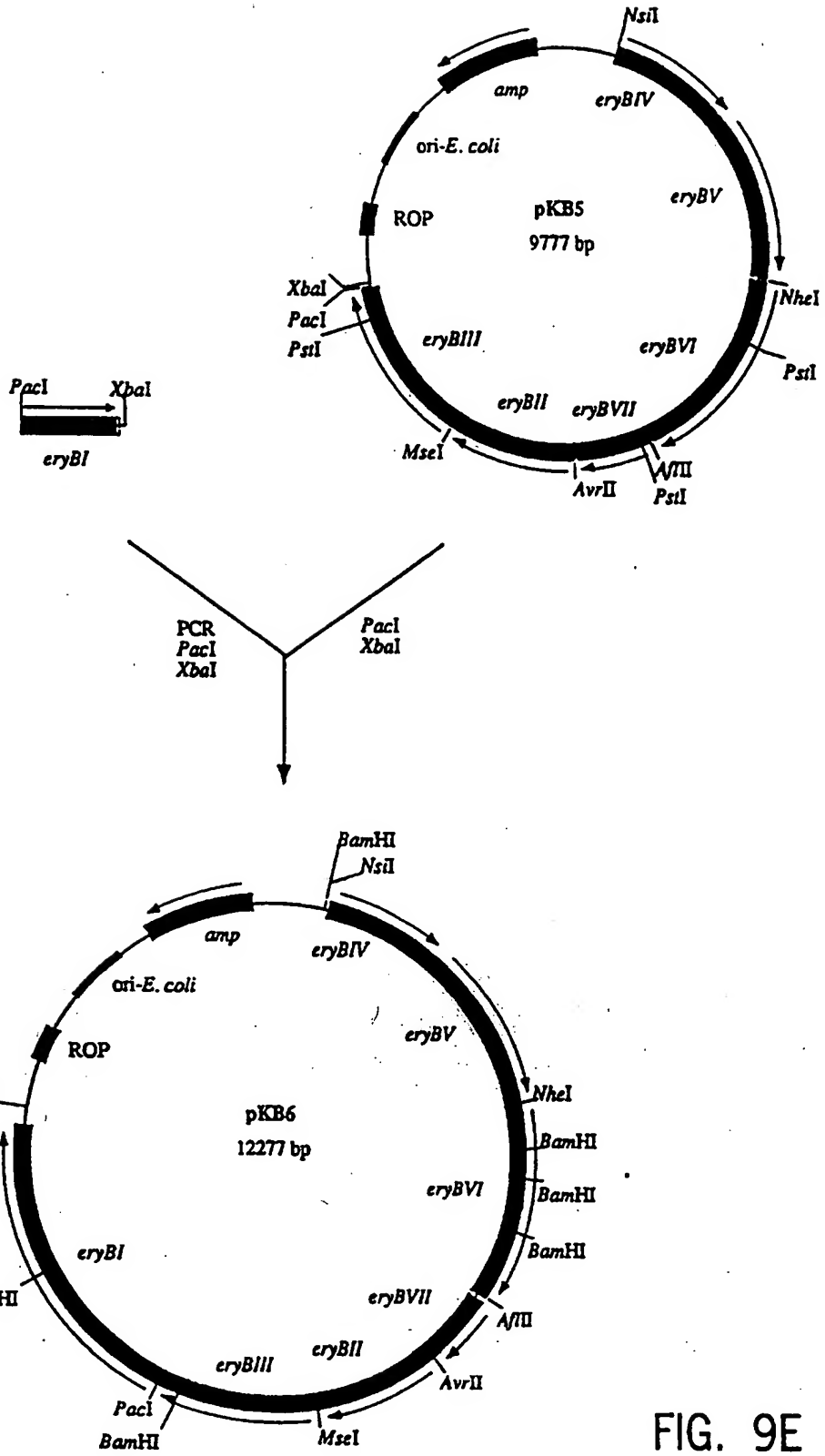
NsiI
amp
eryBIV
ori-E. coli
eryBV
ROP
pKB5
9777 bp
XbaI
PacI
PstI
NheI
eryBIII
eryBVI
PstI
PacI
XbaI
eryBI
eryBII
eryBVII
MseI
AvrII
AflII
PstI
PCR
PacI
XbaI
PacI
XbaI
BamHI
NsiI
amp
eryBIV
ori-E. coli
eryBV
ROP
XbaI
pKB6
12277 bp
NheI
BamHI
eryBVI
BamHI
BamHI
eryBI
BamHI
eryBVII
AflII
eryBII
eryBIII
AvrII
PacI
BamHI
MseI

FIG. 9E

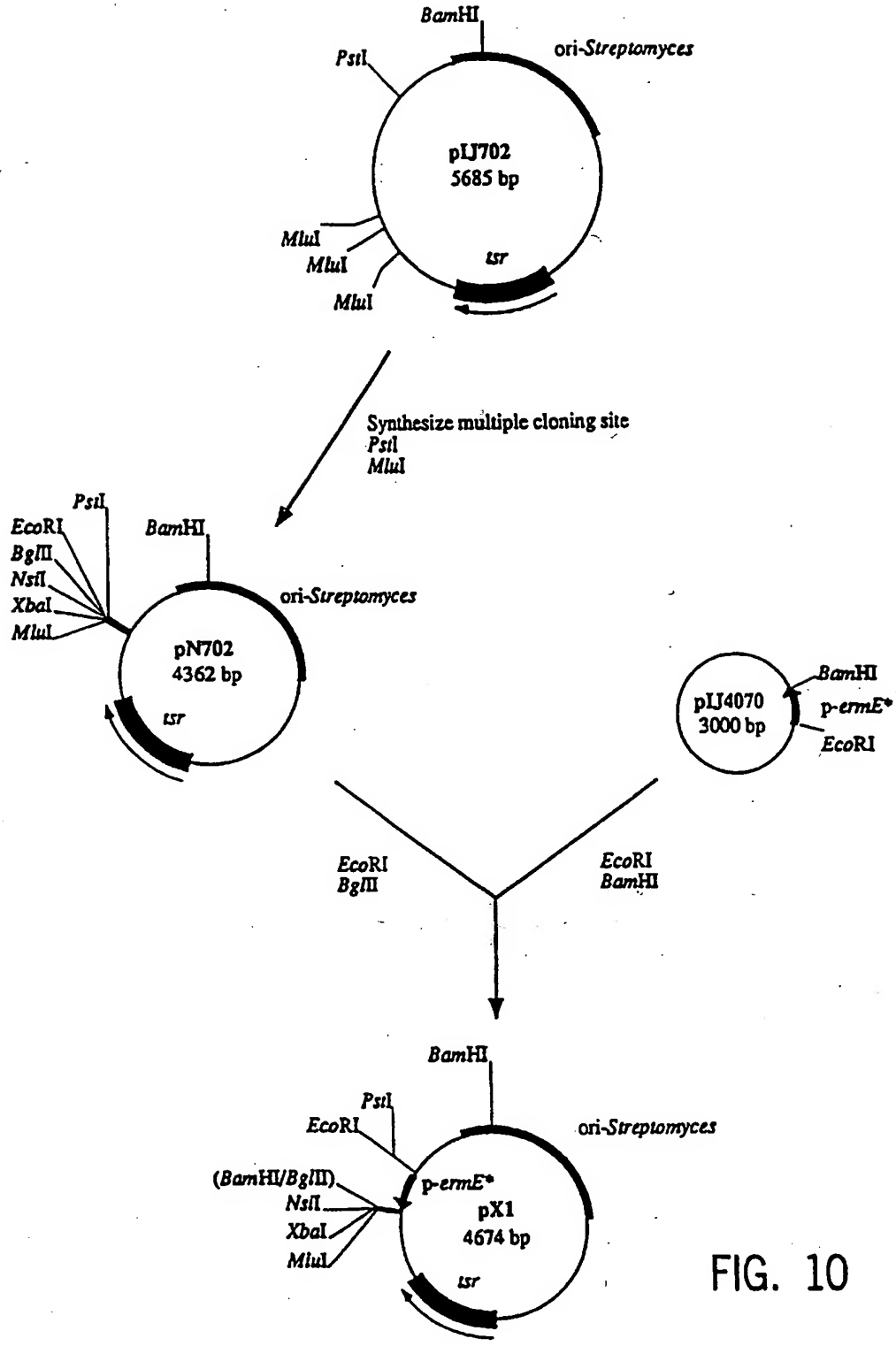
BamHI
PstI
ori-Streptomyces
pIJ702
5685 bp
MluI
MluI
MluI
tsr
Synthesize multiple cloning site
PstI
MluI
PstI
EcoRI
BglII
NsiI
XbaI
MluI
BamHI
ori-Streptomyces
pN702
4362 bp
tsr
pIJ4070
3000 bp
BamHI
p-ermE*
EcoRI
EcoRI
BglII
EcoRI
BamHI
BamHI
PstI
EcoRI
ori-Streptomyces
(BamHI/BglII)
NsiI
XbaI
MluI
p-ermE*
pX1
4674 bp
tsr

FIG. 10

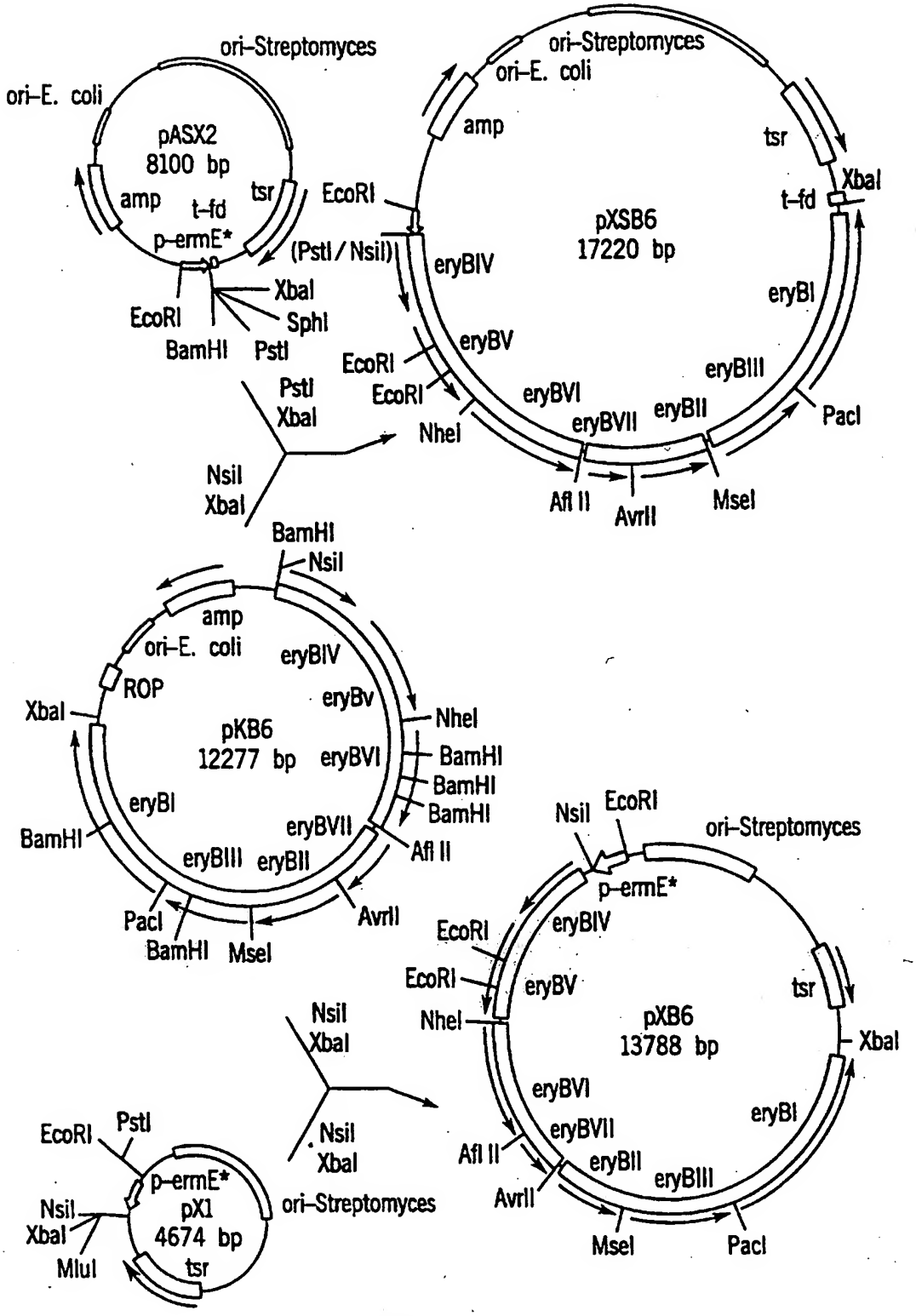
ori-Streptomyces
ori-E. coli
pASX2
8100 bp
amp
tsr
t-fd
p-ermE*
EcoRI
BamHI
PstI
XbaI
SphI
(PstI / NsiI)
pXSB6
17220 bp
eryBIV
eryBV
eryBVI
eryBVII
eryBIII
eryBII
eryBI
NheI
AflII
AvrII
MseI
PacI
NsiI
pKB6
12277 bp
ROP
BamHI
pXB6
13788 bp
pX1
4674 bp
MluI

FIG. 11

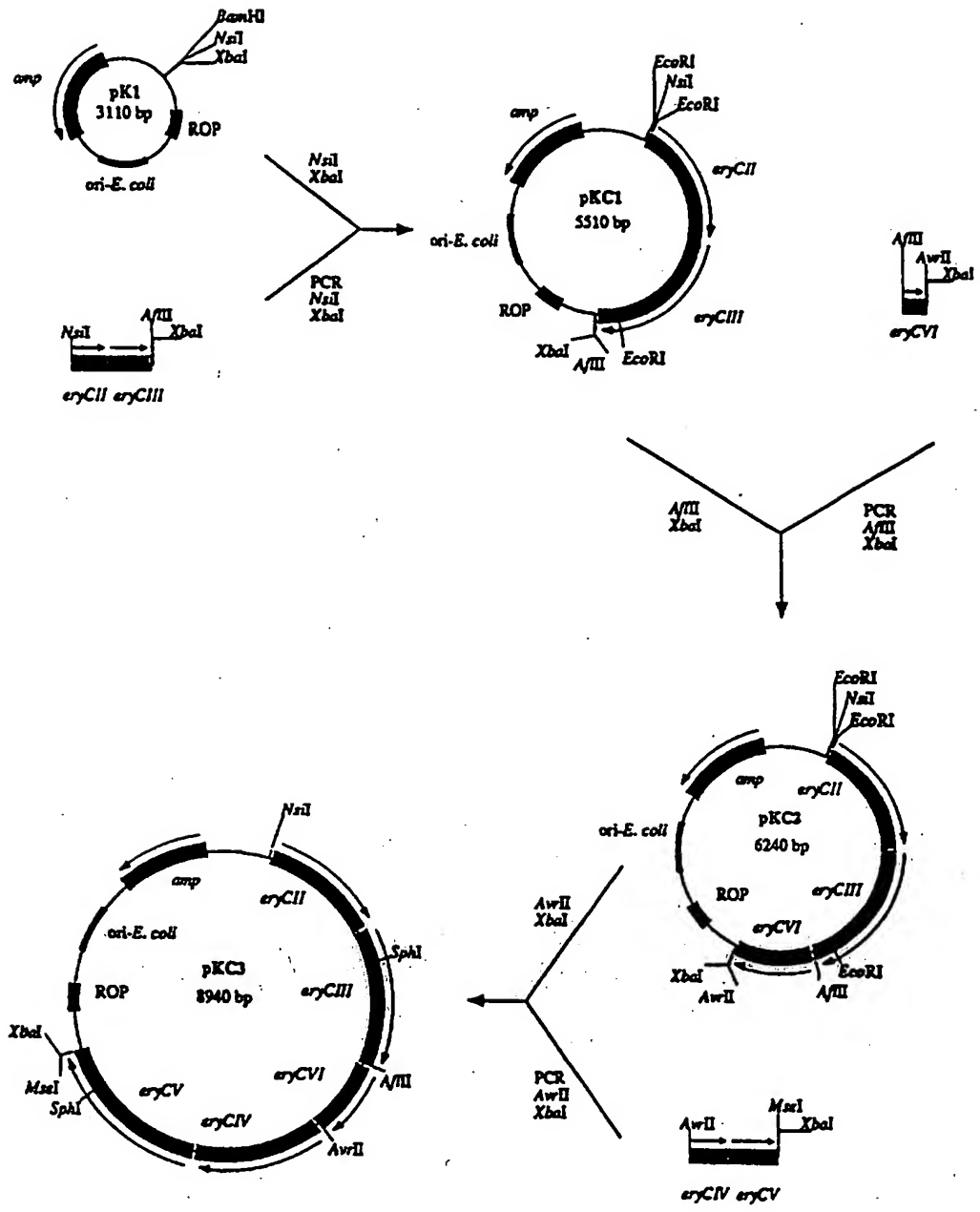
pK1
3110 bp
amp
ori-E. coli
ROP
BamHI
NsiI
XbaI
NsiI
XbaI
PCR
NsiI
XbaI
NsiI
AflII
XbaI
eryCII eryCIII
pKC1
5510 bp
amp
ori-E. coli
ROP
EcoRI
NsiI
EcoRI
eryCII
eryCIII
XbaI
AflII
EcoRI
AflII
AvrII
XbaI
eryCVI
AflII
XbaI
PCR
AflII
XbaI
pKC2
6240 bp
amp
ori-E. coli
ROP
eryCII
eryCIII
eryCVI
EcoRI
NsiI
EcoRI
XbaI
AvrII
AflII
EcoRI
AvrII
XbaI
PCR
AvrII
XbaI
AvrII
MstI
XbaI
eryCIV eryCV
pKC3
8940 bp
amp
ori-E. coli
ROP
eryCII
eryCIII
eryCVI
eryCIV
eryCV
NsiI
SphI
AflII
AvrII
XbaI
MstI
SphI

FIG. 12A

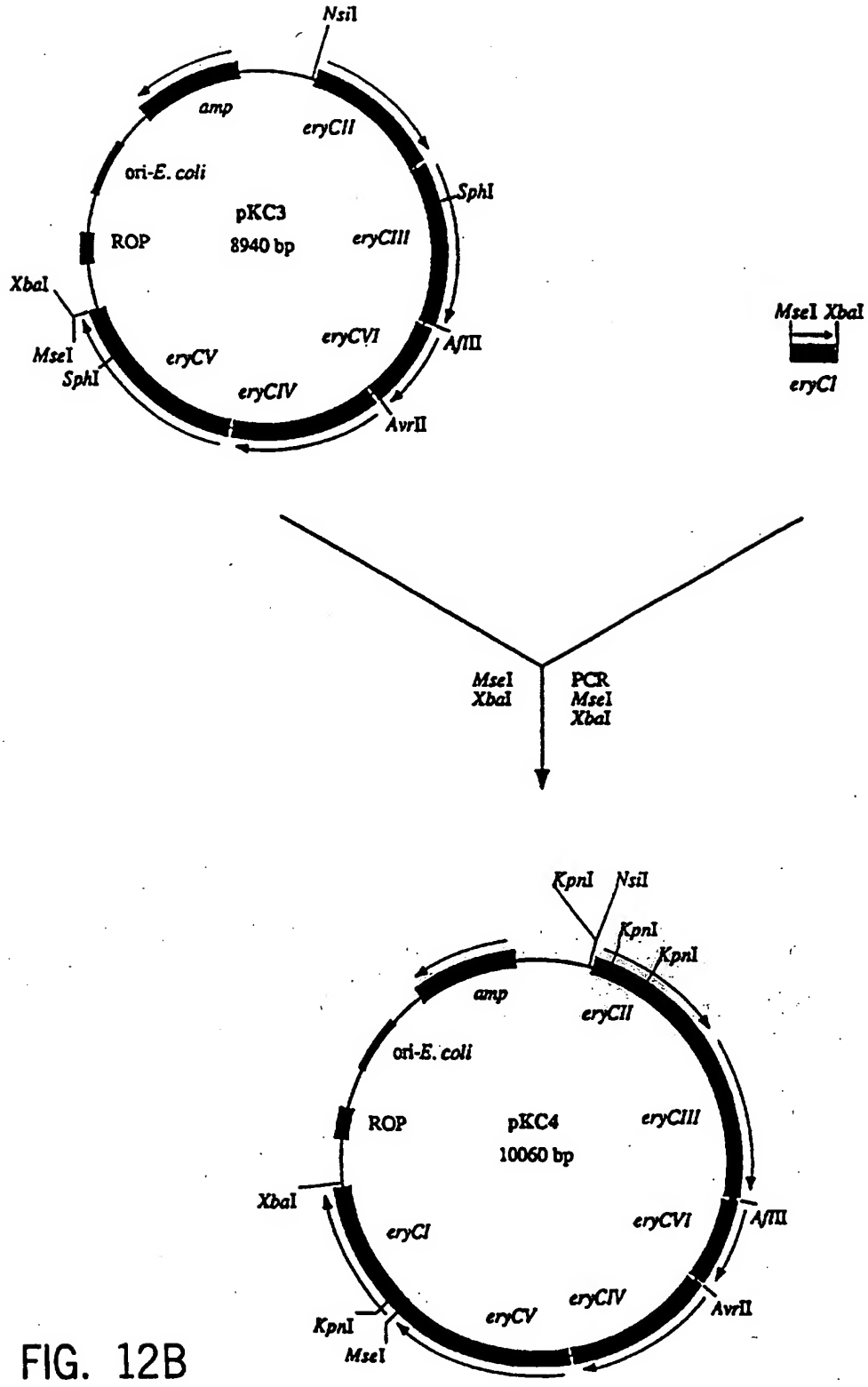
NsiI
amp
eryCII
ori-E. coli
SphI
pKC3
ROP
8940 bp
eryCIII
XbaI
eryCVI
AflII
MseI
SphI
eryCV
eryCIV
AvrII
MseI XbaI
eryCI
MseI
XbaI
PCR
MseI
XbaI
KpnI
NsiI
KpnI
KpnI
amp
eryCII
ori-E. coli
ROP
pKC4
10060 bp
eryCIII
XbaI
eryCI
eryCVI
AflII
eryCIV
eryCV
AvrII
KpnI
MseI

FIG. 12B

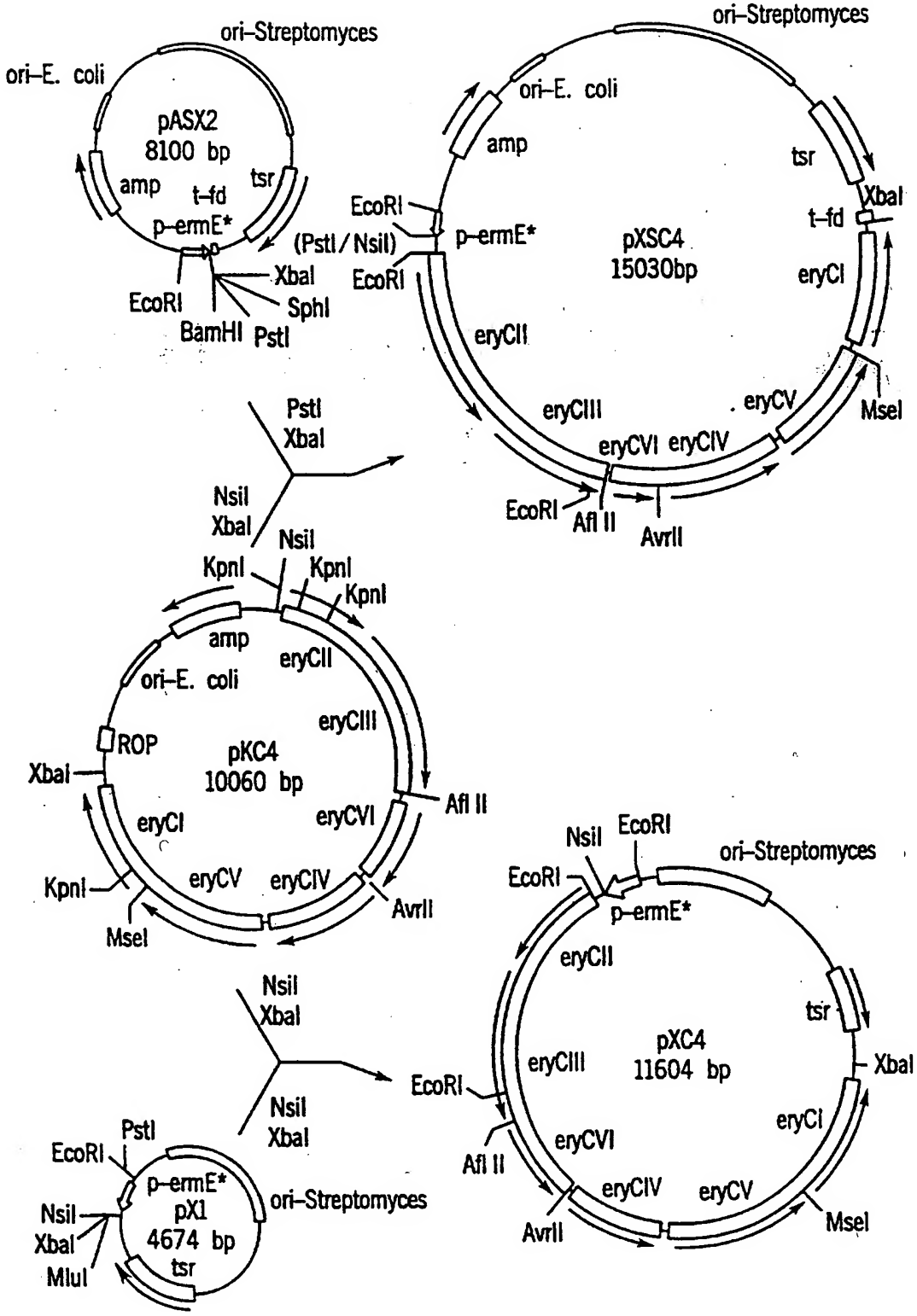
ori–Streptomyces
ori–E. coli
pASX2
8100 bp
amp
tsr
t–fd
p–ermE*
EcoRI
BamHI
PstI
XbaI
SphI
ori–Streptomyces
ori–E. coli
amp
tsr
XbaI
t–fd
EcoRI
p–ermE*
(PstI / NsiI)
EcoRI
pXSC4
15030bp
eryCI
eryCII
eryCIII
eryCV
MseI
eryCVI
eryCIV
EcoRI
AfI II
AvrII
PstI
XbaI
NsiI
XbaI
NsiI
KpnI
KpnI
KpnI
amp
eryCII
ori–E. coli
eryCIII
ROP
pKC4
10060 bp
XbaI
AfI II
eryCI
eryCVI
KpnI
eryCV
eryCIV
AvrII
MseI
NsiI
EcoRI
ori–Streptomyces
EcoRI
p–ermE*
eryCII
tsr
pXC4
11604 bp
eryCIII
XbaI
EcoRI
eryCI
eryCVI
AfI II
eryCIV
eryCV
AvrII
MseI
NsiI
XbaI
NsiI
XbaI
PstI
EcoRI
p–ermE*
ori–Streptomyces
pX1
4674 bp
NsiI
XbaI
MluI
tsr

FIG. 13